Volatility and correlation modelling for equity indices

Dissertation
submitted to the Faculty of Economics,
Business Administration and Information Technology
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Chris Bardgett
from France

approved in July 2014 at the request of

Prof. Dr. Markus Leippold
Prof. Dr. Rama Cont

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 16 July 2014

Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller.

Acknowledgements

I am deeply thankful to Prof. Dr. Markus Leippold for supervising my thesis, making it possible for me to work on various interesting topics, always available for discussion, and caring for a pleasant atmosphere in our team. He also gave me the opportunity to prepare the Financial Engineering lecture with Elise Gourier and teach it with Elise and then Nikola Vasiljević. Furthermore, I am grateful to Prof. Dr. Rama Cont first for accepting my visit at the Imperial College for 6 months in the Maths-Finance department and then for accepting to co-supervise my thesis. I am impressed at all the interesting and diverse ideas he has and thankful for the critical and rigorous feedback he gives. I would also like to thank Prof. Dr. Loriano Mancini (SFI professor) for attending my defense. I really enjoyed his lecture in econometrics and his enthusiasm rubbed off on me, in particular concerning the estimation of market risk neutral distributions.

I am also grateful to Prof. Dr. Christoph Schwab who was available for discussion and gave me feedback for one of my papers. I learned a lot from his vast knowledge of numerical techniques, the first time during my masters at the ETHZ/UZH. I also want to thank my co-author Elise, who has helped me tremendously, giving me feedback and discussing with me long hours on all possible aspects of my PhD.

I feel privileged to have worked on publications together with Markus Leippold, Elise Gourier and Christoph Schwab. I hope I can keep learning from them and work together in the future.

I would also like to thank Prof. Dr. Marc Paolella as program coordinator for his support, encouragement and pleasantness.

I want to thank all the people from my chair: Markus, Elise, Nikola, Lujing, Felix and Meriton for all the good times we had together and to my colleagues in the "old" building of Platte 22 for the nice atmosphere. In particular, I will always remember our team meetings in Zermatt.

Finally, I want to thank Iris Scherwitzl and my family for being encouraging, supportive and listening to me complain about things when they didn't work.

Zürich, January 2014 Chris Bardgett

Contents

Part I

Introduction

Introduction and Summary of Research Results

Modelling the evolution over time of financial returns is one of the most important topics in Quantitative Finance. The essential tasks of portfolio allocation and risk management such as calculating risk measures for portfolios of assets, pricing and hedging accurately derivatives, forecasting, all rely on the ability to design models that capture appropriately the dynamics of asset prices. Because market prices are after all driven by the trading of people and not by a physical law, one can argue that the dynamics of returns is not fixed and may change over time. Before the crash of 1987, the implied volatilities of equity index options were rather flat across strikes (indicating that traders believed in the model of Black and Scholes (1973)) and nowadays index implied volatilities have a strong negative skew. As a consequence of this crash, people became aware that financial returns have heavy tailed distributions and may jump. The choice or design of a model for returns should be guided by the characteristics of the (historical) data, having in mind that there is usually a trade-off between model realism/complexity and tractability. Since the 1987 crisis, a lot of research has been carried out to analyze asset returns and regardless of the asset class considered, returns share common features, referred to as stylized facts[1]. It is now widely documented[2] that the volatility of returns and the correlation between them are time-varying. Many properties of the volatility (or variance) process have been identified[3], such as mean-reversion, negative correlation with asset returns (leverage effect), sharp increases but high persistence (i.e., the possibility that volatility stays low or high for a "long" period of time). To model the dynamics of more than one asset, it is certainly important to model the volatility of each asset but additionally, one has to model the dependence structure of these assets. Similarly to the constant volatility assumption before the 1987 crash, the assumption of constant correlation took a serious hit especially in the last years due to the presence of systemic risk in the economy. In fact, the average implied correlation between the components of the S&P 500 index went from approximately 40% in 2007 to 70% in 2008 and 85% in 2011. The correlation between asset returns share some of the stylized facts of volatility: moves in correlation and returns are negatively correlated, volatility and correlation tend to move together and persistence of correlation. [4]

In the past decades, research in quantitative finance has focused a lot on how to model the dynamics of one asset whether in the econometrics or option pricing literature. The state of the current research

[1] A detailed review of the stylized facts of asset returns can be found in Cont (2001).
[2] Among others, we refer to Engle (1982), Heston (1993), Dupire (1994), Derman and Kani (1994) and Bollerslev, Engle, and Wooldridge (1988), Engle (2000), Chesnay and Jondeau (2001), Moskowitz (2003).
[3] We refer to Engle and Patton (2000), Bakshi, Ju, and Ou-Yang (2006), Bates (1996), Gatheral (2006) among others.
[4] We refer to Chesnay and Jondeau (2001), Skintzi and Refenes (2005), Kaya Boortz (2008).

is that returns are driven by stochastic volatility[5] (with potentially more than one factor[6]), jumps[7] and there is increasingly more evidence towards the existence of jumps in volatility[8]. Interestingly, there is much less literature on modelling the joint dynamics of asset prices and most of the existing research focuses on modelling a stochastic covariance matrix[9] and a few on modelling a stochastic correlation matrix. [10] The reasons seem to be mainly technical. First, capturing the joint dynamics of many assets generally means adding stochastic factors, hence increasing dramatically the computational power needed for estimation. Second, it is theoretically challenging to build a model for which asset correlations evolve randomly and yet the stochastic correlation matrix remains almost surely positive definite over time.

Once the model is chosen, the estimation and calibration[11] of models is also very important and the method used should depend on the application in mind. Choosing different loss functions when calibrating option pricing models will lead to different parameters and therefore different in-sample and out-of-sample results. Christoffersen and Jacobs (2004) show that to achieve the best in-sample fit of option prices, a model should be calibrated to option prices (and not implied volatilities for instance). However, the choice of loss function to achieve the best out-of-sample results is not clear. To infer the model parameters, it is standard practice to calibrate a chosen option pricing model to a short period of time (typically one day) and to re-calibrate the model parameters as soon as the model cannot match the observed options' data anymore. This definitely provides the best in-sample fit of option prices for a given model, however, this estimation method can be seriously questioned when trying to determine whether a model has accurate dynamics or not. If one changes the parameters of the model constantly, one is testing the ability of a model to interpolate option prices (replicate the market implied risk-neutral distributions[12] of the returns) but not the ability of a model to replicate the dynamics of option prices. Regularly re-calibrating models makes it difficult to understand the dynamics of returns, crucial for hedging purposes for instance. Moreover, backing out unobservable processes - such as the volatility and correlation - from calibrations yields estimates for the trajectories of these latent factors that do not follow the dynamics of the model, which in turn makes it difficult to understand well the dynamics of financial returns and option prices via calibration. Many of these issues can be solved to a large extent by using filtering techniques to estimate a model. Using filters,

[5]Given the amount of literature on stochatic volatility models, we only give some references: Engle (1982), Dupire (1994), Derman and Kani (1994), Heston (1993), Bates (1996), Eraker, Johannes, and Polson (2003).
[6]See Cont and da Fonseca (2002), Christoffersen, Heston, and Jacobs (2009), Gruber, Tebaldi, and Trojani (2010).
[7]Barndorff-Nielsen and Shephard (2006), Chernov, Gallant, Ghysels, and Tauchen (2003), Cont and Mancini (2011) provide evidence of jumps in returns using either time series of returns or option price data with short maturities. The list is non exhaustive.
[8]Volatility being intrinsically unobservable, it is more difficult to detect its jumps, however, with the introduction of volatility indices and derivatives (VIX index, VIX options in particular), the drivers of volatility are becoming increasingly identifiable. We refer in particular to Eraker, Johannes, and Polson (2003), Todorov and Tauchen (2011), Kaeck and Alexander (2012), Mencía and Sentana (2013).
[9]We refer in particular to models driven by Wishart process and jump extensions (Gouriéroux, Jasiak, and Sufana (2004), Leippold and Trojani (2008), Cuchiero, Filipovic, Mayerhofer, and Teichmann (2009))
[10]We refer to Kaya Boortz (2008), Driessen, Maenhout, and Vilkov (2009), Ahdida and Alfonsi (2013).
[11]By calibration, we refer to the standard practice of inferring parameters for option pricing models by matching option prices. Calibration is usually done only to one day of option prices and not performed using likelihood criteria, but by minimizing some chosen distance between the model and market prices.
[12]Provided there exists an infinite amount of strikes traded, the knowledge of option prices for all strikes is the same as that of the knowledge of the marginal distribution of the returns under a risk-neutral measure.

one can infer a unique set of parameters for a model as well as the corresponding most likely trajectories for the unobservable processes given returns and options' data.

Finally, to learn about the dynamics of asset prices, the choice of dataset as well as data treatments performed have an impact on the quality of the estimates. At the money options are the most liquid options traded and therefore are the most reliable quotes, however they provide much less information about extreme events than out of the money options. Including more liquid data in the estimation dataset helps provide more reliable estimates for model parameters, but clearly makes the estimation procedure more cumbersome and computationally intensive. Being heavily traded and having many options with high trading volumes, equity indices are the ideal asset to study the dynamics of financial returns. In particular, the S&P 500 index additionally has a volatility index - the VIX - which represents the expected future realized volatility of the S&P 500 returns over the next month. Options on the VIX started trading in 2006 and now have one of the highest trading volume[13] among index options in the world. This makes the S&P 500 index unique with regards to the amount and quality of data available.

This doctoral thesis entitled *Volatility and correlation modelling for equity indices* is composed of three papers addressing some of the important questions raised above. Each paper is presented in a separate chapter as follows.

In the first chapter *Inferring volatility dynamics and risk-premia from the S&P 500 and VIX markets*, we investigate the information contained in S&P 500 returns, VIX levels, S&P 500 and VIX option prices to understand better the dynamics of the S&P 500 index[14]. We develop a rigorous time-series estimation approach using a particle filter and provide an extensive model specification analysis. We find that the S&P 500 and VIX derivatives markets contain conflicting information on variance, especially in times of market stress. Furthermore, jumps and a stochastic level of reversion for the variance help reproduce the tails of returns and variance risk-neutral distributions as well as term structures of volatility smiles. Finally, we observe that they add significant value in representing the variance risk premia accurately.

In the second chapter *A parsimonious stochastic correlation framework to model the joint dynamics of assets*, we introduce a stochastic correlation framework for asset returns each asset following a stochastic volatility model. We show that the system of stochastic differential equations admits a unique strong solution and that the correlation matrix is positive semidefinite over time. Modelling the correlation matrix instead of the covariance matrix provides important computational advantages: the dynamics of each asset can be estimated separately from the dynamics of the correlation matrix making computations less intensive. [15] Furthermore the framework is parsimonious in the number

[13]The average daily trading volume for VIX options was around 400,000 contracts in 2011, about one half of the volume for S&P 500 options.

[14]To extract as much information as possible on the tails of the returns and the volatility distributions, we use a wide range of moneyness for options in both markets.

[15]The choice of modelling the covariance of assets (as in Da Fonseca, Grasselli, and Tebaldi (2007)) instead of correlations implies that the dependence dynamics of assets is mixed together with the univariate dynamics of individual assets. This means that the dynamics of any particular asset depends on all $O(n^2)$ stochastic factors (where n is the number of assets) which is not necessarily intuitive and leads to computationally intensive algorithms.

of stochastic factors which is proportional to the number n of assets, as opposed to quadratic in n. In a numerical experiment, we show that stochastic correlations increases the steepness of the implied volatility smile of index options. To avoid the curse of dimensionality[16] when pricing basket options, we propose two solutions. The first one is to use standard Monte Carlo techniques. The second one is to solve the high dimensional partial differential equation that option prices satisfy using the Quantized Tensor Train representation for large matrices entering in the Finite Difference discretization. This low parametric format for high dimensional tensors makes it possible for the storage cost and computational complexity to grow linearly with the number of assets.

In the third chapter *Model calibration to marginal distributions*, we present an alternative choice of loss function for the calibration of option pricing models and investigate its impact on in and out of sample pricing performance. Given the cross section of options' data, we build market implied risk-neutral distributions (RND) to which we calibrate the model by minimizing a distance criterion between market and model RNDs. We study the advantages and drawbacks of adding distributional assumptions to the options' data for in and out of sample pricing performance, depending on how many strikes are traded. We find that when the number of traded strikes is small, the calibration to the market distributions can lead to smaller out of sample pricing errors but not systematically. On the other hand, when many strikes are traded, even though RND calibration produces in and out of sample prices within the bid-ask spread, calibration to option prices yields smaller errors[17]. We also show that there is no loss of computational efficiency when calibrating to RNDs compared to option prices.

[16]By curse of dimensionality, we refer to the fact that the memory storage needed and number of operations to be computed grows exponentially with the number of assets.
[17]This is in line with the result of Christoffersen and Jacobs (2004).

REFERENCES

Ahdida, A., and A. Alfonsi, 2013, "A mean-reverting SDE on correlation matrices," *Stochastic Processes and their Applications*, 123(4), 1472–1520.

Bakshi, G., N. Ju, and H. Ou-Yang, 2006, "Estimation of Continuous-Time Models with an Application to Equity Volatility Dynamics," *Journal of Financial Economics*, 82, 227–249.

Barndorff-Nielsen, O. E., and N. Shephard, 2006, "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4(1), 1–30.

Bates, D. S., 1996, "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge, 1988, "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96(1), pp. 116–131.

Chernov, M., A. R. Gallant, E. Ghysels, and G. T. Tauchen, 2003, "Alternative Models of Stock Price Dynamics," *Journal of Econometrics*, 116, 225 – 257.

Chesnay, F., and E. Jondeau, 2001, "Does Correlation Between Stock Returns Really Increase During Turbulent Periods?," *Economic Notes*, 30(1), 53–80.

Christoffersen, P., and K. Jacobs, 2004, "The importance of the loss function in option valuation," *Journal of Financial Economics*, 72(2), 291 – 318.

Christoffersen, P. F., S. Heston, and K. Jacobs, 2009, "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well.," *Management Science*, 55(12), 1914–1932.

Cont, R., 2001, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, 1, 223–236.

Cont, R., and J. da Fonseca, 2002, "Dynamics of implied volatility surfaces," *Quantitative Finance*, 2(1), 45–60.

Cont, R., and C. Mancini, 2011, "Nonparametric tests for pathwise properties of semimartingales," *Bernoulli*, 17(2), 781–813.

Cuchiero, C., D. Filipovic, E. Mayerhofer, and J. Teichmann, 2009, "Affine Processes on Positive Semidefinite Matrices," working paper, ETH Zurich, manuscript.

Da Fonseca, J., M. Grasselli, and C. Tebaldi, 2007, "Option pricing when correlations are stochastic: an analytical framework," *Review of Derivatives Research*, 10(2), 151–180.

Derman, E., and I. Kani, 1994, "Riding on the Smile," *RISK*, 7, 32–39.

Driessen, J., P. J. Maenhout, and G. Vilkov, 2009, "The Price of Correlation Risk: Evidence from Equity Options," *Journal of Finance*, 64(3), 1377–1406.

Dupire, B., 1994, "Pricing with a Smile," *RISK*, 7, 18–20.

Engle, R., 2000, "Dynamic Conditional Correlation - A Simple Class Of Multivariate Garch Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F., 1982, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

Engle, R. F., and A. J. Patton, 2000, "What Good is a Volatility Model," *Quantitative Finance*, pp. 237–245.

Eraker, B., M. S. Johannes, and N. Polson, 2003, "The Impact of Jumps in Equity Index Volatility and Returns," *Journal of Finance*, 58, 1269–1300.

Gatheral, J., 2006, *The volatility surface: a practitioner's guide*, vol. 357. Wiley.

Gouriéroux, C., J. Jasiak, and R. Sufana, 2004, "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," working paper, DP, University of Toronto.

Gruber, P., C. Tebaldi, and F. Trojani, 2010, "Three Make a Dynamic Smile - Unspanned Skewness and Interacting Volatility Components in Option Valuation," .

Heston, S., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

Kaeck, A., and C. Alexander, 2012, "Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions," *Journal of Banking & Finance*, 36(11), 3110 – 3121.

Kaya Boortz, C., 2008, "Modelling correlation risk.," *Diplomarbeit, preprint.*

Leippold, M., and F. Trojani, 2008, "Asset Pricing with Matrix Jump Diffusions," Working paper.

Mencía, J., and E. Sentana, 2013, "Valuation of VIX derivatives," *Journal of Financial Economics*, 108(2), 367 – 391.

Moskowitz, T. J., 2003, "An Analysis of Covariance Risk and Pricing Anomalies," *Review of Financial Studies*, 16(2), 417–457.

Skintzi, V. D., and A.-P. N. Refenes, 2005, "Implied correlation index: A new measure of diversification," *Journal of Futures Markets*, 25(2), 171–197.

Todorov, V., and G. Tauchen, 2011, "Volatility Jumps," *Journal of Business and Economic Statistics*, 29, 356–371.

Part II

Research Papers

Inferring volatility dynamics and risk-premia from the S&P 500 and VIX markets

Chris Bardgett, Elise Gourier and Markus Leippold

Abstract

This paper investigates the information contained in S&P 500 returns, VIX levels, S&P 500 and VIX option prices. We develop a rigorous time-series estimation approach and provide an extensive model specification analysis. We find that the S&P 500 and VIX derivatives markets contain conflicting information on variance, especially in times of market stress. Furthermore, jumps and a stochastic level of reversion for the variance help reproduce the tails of returns and variance risk-neutral distributions as well as term structures of volatility smiles. Finally, we observe that they add significant value in representing the variance risk premia accurately.

KEYWORDS: S&P 500 and VIX joint modeling, option pricing, particle filter, dynamics of volatility.

1 Introduction

One of the central questions addressed by research in empirical option pricing is the determination of asset returns dynamics. Ideally, in addition to reproducing asset prices and prices of derivatives traded on a given day on the market, they should also recreate the joint evolution of these prices over time. Under the historical measure $\mathbb{P}$, the time series of asset returns provides valuable information on the main characteristics of returns dynamics (assuming stationarity and ergodicity). On the other hand option prices on this asset help specify its dynamics under the risk neutral measure $\mathbb{Q}$. Indeed, the result of Breeden and Litzenberger (1978) states that the observation of vanilla option prices with maturity T for a continuum of strikes entirely determines the $\mathbb{Q}$ distribution of this asset at the future time T. Even though we do not observe prices for arbitrary strikes in practice, the S&P 500 index has many strikes traded liquidly. Whereas a large part of the literature on asset pricing either focuses on the time-series properties of returns under the historical measure or proposes models to accurately capture the stylized facts of option prices, the study of the link between both measures has recently captured more attention. The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is achieved through an appropriate specification of risk premia, which can be interpreted as compensations for the risks that investors take when trading an asset. While there is a large amount of research articles and surveys on the equity risk premium, the study of the variance dynamics and variance risk premium is more recent. Prominent examples include Bates (1996, 2000, 2003), Chernov and Ghysels (2000), Jackwerth (2000), Pan (2002), Jones (2003), Eraker (2004), Aït-Sahalia and Kimmel (2007), Broadie, Chernov, and Johannes (2007), Carr and Wu (2009), Egloff, Leippold, and Wu (2010), Todorov (2010), Wu (2011) and Aït-Sahalia, Karaman, and Mancini (2012). However, the components of risk premia, in particular when jumps are involved, are usually found hard to estimate and statistically insignificant with daily data. One reason for this is that the estimation of risk premia requires a large amount of returns and options data and therefore powerful computational tools to extract the relevant information. In fact, because of this computational burden, most research does not make use of the whole cross section of options and considerably reduces the amount of information. As available sources of information have grown tremendously since the introduction of the volatility index VIX as well as VIX derivatives, the need for efficient computational algorithms that can exploit all this information arises.

The VIX index has been constructed to approximate non-parametrically the expected future realized volatility of the S&P 500 returns over the next 30 days. The index is not directly tradeable but it is possible to trade VIX futures and options. Options started trading in 2006 and have been a fast growing business ever since, nowadays representing a much larger market than VIX futures. By definition, the VIX index is linked to the dynamics of the S&P 500 index returns and this makes VIX and S&P 500 options both ideal to infer these dynamics. Fortunately, the S&P 500 and VIX options markets are among the most liquid worldwide with a daily average volume of 783,768 and 391,992 contracts traded per day in 2011, and therefore represent a trustworthy source of information. Including more information on volatility and its evolution over time is essential to better specify and understand the dynamics of volatility. Subsequently, we will use interchangeably S&P 500 and SPX, which is its ticker symbol.

Our paper mainly distinguishes itself from the literature because it uses joint datasets in a unique manner. We use time series of SPX and VIX derivatives in order to make inference on the dynamics of returns and volatility as well as risk premia specification. Since SPX and VIX options both provide information on the same volatility and jump processes, any model which can price one derivatives market but not the other is bound to have strongly misspecified dynamics and risk premia. Incorporating both the SPX and VIX options therefore gives us a chance to fine-tune the specification of volatility dynamics and study risk premia using a larger source of information than previously done. Furthermore, we base our analysis on a time series of cross sectional data of options with a wide range of moneynesses and maturities. We emphasize that we have kept liquid deep out-of-the money options in our dataset in contrast to most of the literature, so as to keep valuable information on the tails of S&P 500 returns and VIX levels. These options are essential in the estimation of the jump structure of their underlying under the risk-neutral measure, i.e., of jumps in the S&P 500 and VIX indices. Indeed, the steepness of the S&P 500 smile and the high volatilities for short maturity puts are considered to be a strong indication of jumps in the returns. Similarly, the positive skewness of VIX implied volatilities and high volatilities of deep out-of-the-money calls suggest the presence of positive jumps in the VIX. Therefore, the cross section of options is required to infer possible jumps under the risk-neutral measure and justifies our choice to use the whole cross section of S&P 500 and VIX options.

Using a dataset containing S&P 500 index levels, VIX values, and prices of options on both markets, we make the following contributions to the empirical option pricing literature.

First, we analyze and compare the information contained in the S&P 500 and VIX markets. We find that when the market is calm, options do not provide more information on the dynamics of volatility than the underlying S&P 500 returns and VIX levels. However, during market turmoil our results indicate that the information contained in S&P 500 options conflicts with the one contained in the underlying index levels or in VIX options. Furthermore, we find that options on both markets provide valuable and complementary information on the level of reversion of the S&P 500 returns' variance as well as on jumps' arrival times and magnitude. These findings are further supported by a thorough analysis of how models perform in the pricing of options they were not estimated to, during the in- and out-of-sample periods. We find that pricing options which are not accounted for in the estimation procedure leads to severe mispricing, and conclude that on the one side, VIX levels do not span the information contained in S&P 500 options, and on the other side, S&P 500 options do not span the information contained in VIX options and vice versa. It is crucial to be aware of this lack of market integration when pricing, risk managing or hedging positions on one market with options on the other one.

Second, we perform an extensive model specification analysis. We detail and explain the role of the different features of the model in explaining option prices and risk-neutral distributions of returns and of their variance process. We model the S&P 500 returns using the affine framework of Duffie, Pan, and Singleton (2000). This structure allows us to price S&P 500 and VIX derivatives in semi-closed form and is essential to carry out the analysis of returns and volatility dynamics using such a large

dataset of options. However, we point out and reduce the limitations of one-factor affine models by advocating a stochastic level of reversion in the volatility dynamics. The flexibility of this model makes it possible to investigate how many factors are needed to reproduce the times series features of the data, and whether jumps should be incorporated. Up to now and to our knowledge, extracting information from both SPX and VIX derivatives markets has not been done and therefore provides new valuable insight into the dynamics of asset returns and volatility. Based on likelihood criteria as well as statistical tests of pricing errors, we find that jumps in the returns and in the variance process are needed to jointly represent the index levels and derivatives prices on both markets. Introducing a stochastic level of reversion for the variance (also known as stochastic central tendency) helps to better represent the tails of the returns' distribution as well as the term structure of S&P 500 and VIX options prices. Furthermore, jumps allow to better represent the right tail of the variance distribution as well as short-maturity options.

Third, estimating the dynamics using such an extremely large dataset of options on the two markets and for a long time series requires computationally efficient techniques that can easily deal with the complicated features of the model, in particular state-dependent jumps. To achieve this goal, we extend the Fourier Cosine method introduced by Fang and Oosterlee (2008) for S&P 500 options to price VIX options and adapt the Auxiliary Particle Filter of Pitt and Shephard (1999) to filter out unobservable processes over time and their jumps. Accordingly, we provide a extensive toolkit for inference and diagnostics of 3-factor affine option pricing models given indices and options data from the SPX and VIX markets. Sequential Monte-Carlo techniques have recently increased in popularity and have been used to estimate models, but most papers using this tool restrict their options dataset to near at-the-money options and to the extent of our knowledge, none have used SPX and VIX derivatives jointly.

Our fourth contribution is the thorough analysis of the equity and variance risk premia. We want to stress that our estimation methodology consists in a single step using the times series of indices and options together, i.e., we jointly estimate the parameters that determine the dynamics of returns under the historical and risk-neutral probability measures. This approach increases the computational complexity but ensures a consistent estimation of the historical and pricing measures, which is essential to estimate reliable risk premia. Although this analysis is model dependent, the flexibility of our model makes it possible to obtain risk premia which are economically sensible and reveals some interesting characteristics. In particular, we find that the integrated equity risk premium is mainly determined by the diffusive shocks in returns, and only marginally affected by rare jumps. Furthermore, the stochastic central tendency only has a minor impact on it, disregarding the maturity considered. Conversely, the variance risk premium is very sensitive to jumps in particular when the maturity is near, since a large movement in the variance process has an immediate negative impact on the payoff of a short-term variance swap. The stochastic central tendency plays a significant role in both continuous and discontinuous parts, especially in calm markets. It is a more persistent factor than v, which represents long-term expectations of investors, and as such the contribution of v takes over in times of market turmoil.

The paper is organized as follows. In Section 2, we explain how our paper fits into the existing literature. We then conduct a preliminary data analysis in Section 3, and highlight some differences between the S&P 500 and the VIX options markets. In Section 4 we present the two-factor affine jump-diffusion model that we use later in the estimation. We describe the risk premium specification and derive the expression for the VIX squared as well as the pricing formula for VIX and S&P 500 options. In Section 5 we discuss the joint estimation to one single day of data as well as the Auxiliary Particle Filter that we use to calibrate the model to a time-series of cross-sectional options data. Finally, in Sections 6 and 7 we summarize the results of the daily and time-series estimations and present our findings. Section 8 concludes.

2 Related literature

Our work builds on an extensive body of research that analyzes which features are needed for a model to provide a realistic representation of equity underlying and derivatives prices. While the end of the twentieth century has been characterized by a fast growing literature on equity option pricing, the financial crisis has recently drawn more attention to the need to better understand and model equity volatility. The volatility index VIX has been introduced in 1993 and is formally defined in the white paper of the CBOE (2009). In practice it is calculated using a combination of S&P 500 options with maturities adjacent to 30 days. Intuitively, the VIX squared is close to the 30-day expected future realized variance and therefore its value should be close to the 30 day-variance swap on the S&P 500 returns. Demeterfi, Derman, Kamal, and Zou (1999) showed that variance swaps can be partially hedged (and therefore priced) using a combination of vanilla options and this is where the formal definition of the VIX comes from. Since the introduction of the VIX and its derivatives (from 2004 onwards), the direct modeling of volatility and the pricing of its derivatives has been the focus of numerous papers. We refer among others to Whaley (1993), Grünbichler and Longstaff (1996), Detemple and Osakwe (2000), Sepp (2008a,b), Bergomi (2009), Lian and Zhu (2011), Drimus and Farkas (2013) and Mencía and Sentana (2013). An important conclusion of this literature is that sharp increases in the variance dynamics are necessary to reproduce the positive skewness of VIX options' implied volatilities. In particular, many articles point out that this could be achieved by having positive jumps in the variance. In particular, Christoffersen, Jacobs, and Mimouni (2010) demonstrate via Q-Q plots that using a square-root model without jumps for the variance is not in line with empirical properties of the data. Using realized variance taken at high frequency as a proxy for the integrated variance, they show that the empirical realized volatility is not Gaussian as the continuous square-root model posits. Jumps allow the distribution of the integrated volatility to be fatter-tailed and therefore represent better the data. Todorov (2010) and Todorov and Tauchen (2011) test for jumps in the VIX index and find strong evidence supporting this assumption. They also test for co-jumps in S&P 500 returns and in the VIX and find striking evidence for them. He finally finds that 63% of the co-jump variation in the sample studied is due to the combination of negative jumps in the returns and positive jumps in the volatility. Jacod and Todorov (2010) develop further statistical tests which indicate that most stock market jumps are associated with volatility jumps. Eraker, Johannes,

and Polson (2003) show that using jumps in the volatility process significantly improves the fit of returns. Finally, as mentioned in Eraker (2004), continuous volatility or variance processes are not able to explain the unusually large volatility before and after the crash of 1987. The specification of jumps is furthermore of importance. Bates (1996), Pan (2002) and Eraker (2004) argue in favor of using state-dependent jumps in returns, which is intuitively appealing as jumps tend to occur more frequently when volatility increases. Using variance swaps, Aït-Sahalia, Karaman, and Mancini (2012) found that the state dependent intensity of jumps was a desirable model feature. However, evidence supporting this choice is mixed. Indeed, Bates (2000) finds that state dependent intensities lead to strong misspecification and Eraker (2004) finds that it does not significantly improve the option prices fit. Broadie, Chernov, and Johannes (2007) and Johannes, Polson, and Stroud (2009) use a constant intensity of jumps.

Another concern of volatility modeling relates to the number of factors that should be used. While adding an additional factor to the Heston model increases the complexity, it has indeed been shown that two factors are needed to provide an accurate description of the volatility dynamics (see, e.g., Andersen, Benzoni, and Lund (2002), Alizadeh, Brandt, and Diebold (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Todorov (2010), Kaeck and Alexander (2012), Bates (2012) and Mencía and Sentana (2013)).

Several papers have been published in the last years aiming to reconcile the cross-sectional information of the S&P 500 and the VIX derivatives markets by modeling them jointly. Gatheral (2008) pointed out first that even though the Heston model performs fairly well to price S&P 500 options, it totally fails to price VIX options. Figure 4 shows that modeling the instantaneous volatility as a square root process leads to a VIX smile decreasing with moneyness, which is the opposite of what is observed in practice. Therefore the volatility density implied by VIX options has more mass at high volatility and less mass at lower volatility levels than the Chi-Square density of the Heston model. Some studies are going in the direction of non-affine models (e.g., Jones (2003), Aït-Sahalia and Kimmel (2007), Christoffersen, Jacobs, and Mimouni (2010), Ferriani and Pastorello (2012), Durham (2012), Kaeck and Alexander (2012)). However tractability remains an issue that is of crucial importance when it comes to calibrating a model to a long time series containing hundreds of options each day.

Among the recent papers that attempted to reproduce simultaneously the smiles of volatility of S&P 500 and VIX options are Chung, Tsai, Wang, and Wenig (2011), Cont and Kokholm (2011), Song and Xiu (2012), Papanicolaou and Sircar (2012) and Bayer, Gatheral, and Karlsmark (2013). We build on this literature by considering extensions of the Heston model that remain in the affine framework, but add more flexibility to the specifications used in the above mentioned papers. Our model is a special case of the general affine framework developed by Duffie, Pan, and Singleton (2000) but includes as sub-cases the usual extensions of the Heston model encountered in the literature, for example Bates (2000), Eraker (2004) and Sepp (2008a).

However most if not all of the papers that consider S&P 500 and VIX options in their calibration exercise have restricted their analysis to a static one-day estimation. Therefore the estimated parameters might exhibit large variations when calibrating the model to different dates. Lindström, Ströjby,

Brodén, Wiktorsson, and Holst (2008) show that the estimated parameters are not stable over time and therefore cannot be used to infer time series properties of returns and risk premia. In the last decade, powerful algorithms have been developed to estimate non-linear models with non-Gaussian innovations in a time-consistent manner.

Time-consistent estimation methods have been used so far to calibrate models to index returns and options. For example, Pan (2002) uses a tailored version of the Generalized Methods of Moments to estimate the Bates model using a time series of S&P 500 and options (two per day). Eraker (2004) relies on Markov Chain Monte Carlo methods to estimate risk premia for jumps in returns and volatility also using returns and options (around three per day). Broadie, Chernov, and Johannes (2007) were the first to consider the whole cross section of option prices on the S&P 500. To reduce the computational burden, they fix some of the parameters by taking values from previous estimations of the time series of returns and minimize a least square type distance between market and model option implied volatilities. They find that the time series provided evidence that volatility jumps, which coincides with the literature that appeared later on VIX option pricing. With a particle filter, Johannes, Polson, and Stroud (2009) investigate whether the time-series of returns of the S&P 500 are consistent with information embedded in option prices. Their options sample is limited to one option per day. They find some inconsistencies that they attribute to either a wrong specification of risk premia or a lack of flexibility of the model. They conclude that their results might be explained by the introduction of a time varying level of reversion for the volatility. Christoffersen, Jacobs, and Mimouni (2010) apply a Maximum Likelihood Importance Sampling technique on returns and a separate Non-linear Least-Squares Important Sampling estimation to option prices to compare the accuracy of models in reproducing returns and option prices. However, as underlined in Ferriani and Pastorello (2012), most papers filtering information from option prices rely on one option per day or a very limited set of options. This is computationally less intensive but ignores a large part of the information present on the market. Ferriani and Pastorello (2012) have used part of the cross section of options and the time series of log-returns in the filtering problem. They do not consider jumps in the volatility but study different non-affine models. They conclude that significant improvement could be brought into these models by incorporating jumps or regime switching in the volatility dynamics. Finally, in a working paper Duan and Yeh (2011) use a filter on the S&P 500 returns together with the VIX index to infer the dynamics of returns and volatility. However, they do not use options data making it impossible to estimate risk premia.

3 Preliminary data analysis

In Figure 1, we plot the joint evolution of the S&P 500 and the VIX index. Their movements are highly negatively correlated, which explains the use of instruments on the VIX to hedge part of the equity risk of a portfolio. Table 1 displays the first four moments of the S&P 500 returns and VIX index levels, over two periods of time. The first period starts in March 2006 and ends November 2008, i.e., it spans the pre-crisis period as well as the beginning of the crisis. The second period begins in December 2008 and lasts until October 2010. S&P 500 log-returns exhibit negative skewness during

the second period considered, and a high kurtosis over both periods, suggesting the presence of rare and large movements. The VIX index exhibits a large skewness and kurtosis in the first period, but in the second period the statistics suggest that the movements are more symmetric, centered about a higher value (29% instead of 20% in the first period).

[Insert Figure 1 here]

[Insert Table 1 here]

We consider closing prices of European options on the S&P 500 from March 1, 2006 to October 29, 2010. The data was obtained from OptionMetrics. The time period of our dataset is restricted by the fact that options on the VIX were introduced in 2006. We also use a dataset of VIX options closing prices on the same time period coming from the data provider DeltaNeutral. This time series includes periods of calm and periods of crisis with extreme events, especially relevant to estimate the presence and magnitude of jumps. In particular, during the financial crisis that started at the beginning of 2007, the VIX index was at its highest peak since its launch.

Both the S&P 500 and VIX options dataset are treated following usual procedures (see Aït-Sahalia and Lo (1998)). In particular, we only consider options with maturity between one week and one year and delete options quotes that where not traded on a given date. We follow two main steps. First, we delete all in-the-money (ITM) options since they are illiquid compared to out-of-the-money (OTM) options. Second, we infer from highly liquid options the Futures price using the at-the-money (ATM) put-call parity. This avoids two issues: Making predictions on future dividends, and using Futures closing prices which are not synchronized with the option closing prices. Hence, we consider that the underlying of the options is the index Futures and not the index itself. At the end, we only work with liquid OTM options for the S&P 500 market and only with liquid call options for the VIX market. Indeed, in the case where the VIX ITM call is not liquid, we use the put-call parity to infer a liquid VIX ITM call from a more liquid VIX OTM put.

These adjustments leave a total of 383,286 OTM S&P 500 options and a total of 43,775 call options on the VIX. This implies a daily average of 327 S&P 500 options and 37 VIX options. The number of S&P 500 options in our dataset on a given date increases with time with around 170 options at the beginning of the dataset and around 450 options at the end. For VIX options, the number increases substantially, with around 5 options per day at the beginning and around 70 options per day at the end. At the beginning of the sample, there are one or two short maturities (below 6 months) available for VIX options and around 6 maturities for S&P 500 options with approximately 40 options per maturity slice. At the end of the sample, VIX options have around 5 short maturities (less than 6 months) with a bit more than 10 options trading per maturity. For S&P 500 options, around ten maturities are available per day with around 60 options for one-month maturities and 40 options for the one-year slice. The low number of VIX options compared to the number of S&P 500 options first comes from the fact that the VIX options market started in 2006 and therefore that the overall volume traded is lower but also from the fact that less maturities and less strikes are traded. At the end of

our sample, the total VIX options volume per day is about half the total volume of S&P 500 options traded but much fewer strikes are traded for VIX options.

To calculate the implied volatilities of VIX options, the true underlying is the VIX Futures value. This can intuitively be explained by the fact that a call option at time t with maturity T is an option on volatility on the time interval $[T, T + 30d]$, where $30d$ stands for 30 days. The value VIX_t at time t is related to volatility on the time interval $[t, t + 30d]$ which might not overlap at all with $[T, T + 30d]$. On the contrary, a Futures on the VIX with maturity T is based on the volatility on the time interval $[T, T + 30d]$. This remark is important because traded VIX option prices do not satisfy no-arbitrage relations with respect to the VIX index, but rather with respect to the VIX Futures value. In particular, calculating implied volatilities assuming that the underlying is the VIX might lead to volatilities equal to zero, or which simply do not exist. For this reason all implied volatilities are calculated with respect to the Futures price of the VIX. The same is done for S&P 500 options as it eliminates the need to make predictions on futures dividends.

Even though the S&P 500 and VIX markets are related, we want to emphasize that VIX options behave in a completely different way than S&P 500 options. First, S&P 500 and VIX derivatives with the same maturity contain different information. On the one hand, an S&P 500 option with maturity T contains information about the future S&P 500 index level at time T and therefore about the S&P 500 volatility up to T. On the other hand, a VIX option with maturity T embeds information about the VIX at time T and therefore about the S&P 500 volatility between T and $T + 30$ days. Second, the implied volatility smiles backed out from S&P 500 and VIX option prices have very different shapes. Figure 2 displays the S&P 500 and VIX smiles depending on different states of the economy. The implied volatilities (IVs) are computed using the Black-Scholes formula, i.e., backing out the standard deviation of a log-normal distribution for the S&P 500 index (respectively for the VIX index) that are implied by their respective option prices. The VIX IVs are in general substantially higher - ranging from 40% to 200%, with an average IV of around 75% (see Table 2) - than S&P 500 IVs (average IV of around 23%). The implied volatilities are negatively skewed for S&P 500 options, generally decreasing with moneyness as risk-averse investors require a premium for negative states of the economy. In contrast, VIX implied volatilities are positively skewed and increase with moneyness, which can intuitively be explained by the fact that negative returns are often observed together with a rise of volatility (the so-called leverage effect) also corresponding to turbulent states of the economy.

[Insert Figure 2 here]

[Insert Table 2 here]

The difference between these markets is also reflected by other indicators such as the put-call trading ratio: Almost twice as many puts as calls are traded daily in the S&P 500 options market but the situation is reversed in the VIX market where the amount of calls traded daily is almost double that of the puts. In fact, one can additionally see in Figure 2 that the log-moneynesses traded for S&P 500 options are mostly negative (which corresponds to out-of-the-money put options) and often positive for VIX options (out-of-the-money calls).

Figure 3 represents the expected forward log-returns of the underlying S&P 500 index returns from March 1st, 2006 to October 29th, 2010 as implied by prices of S&P 500 options with maturity 1 month. We use the method described in Bakshi, Kapadia, and Madan (2003) to calculate the moments implied by option prices. The expected forward returns illustrates the variety of market situations that our time series includes. They were almost constant until the end of 2007, equal to a positive value and thus indicating that market participants were expecting a stable income from investing in the index. But from the end of 2007 they exhibit more variation and seem to mean-revert around a negative trend. Suddenly, following the bail out of Lehman Brothers in September 2008, expected forward returns drop and reach -2% beginning of October 2008. Then they gradually come back and stabilize in mid-2009 around a slightly negative level close to -0.2%. In 2010, the sudden increase in the VIX index coincides with a peak of the expected forward returns reaching about -0.8%. Both the VIX index and expected forward returns as implied by S&P 500 options indicate market expectations over the next month as reflected in index option prices. However volatility provides information on returns through the leverage effect, while the implied expected forward returns are a direct measure of how investors expect returns to behave. They are much more stable in quiet periods and better reflect the different market situations that compose our time-series and that we aim to reproduce with a model.

[Insert Figure 3 here]

4 Model and option pricing

In this section we present the two-factor affine model that we use, and the properties that make it attractive.

4.1 Model specification

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual assumptions, where $\mathbb{P}$ denotes the historical measure. We consider a risk-neutral measure $\mathbb{Q}$[18] equivalent to $\mathbb{P}$ and denote by $(F_t)_{t\geq 0}$ the forward price[19] of the S&P 500 index and by $Y = (Y_t)_{t\geq 0} = (\log(F_t))_{t\geq 0}$ the returns. The dynamics of Y under $\mathbb{Q}$ are specified as follows:

$$dY_t = [-\lambda^{Yv}(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{Q})}(1,0,0) - 1) - \frac{1}{2}v_{t^-}]dt + \sqrt{v_{t^-}}dW_t^{Y(\mathbb{Q})} + dJ_t^{Y(\mathbb{Q})} \quad (1)$$

$$dv_t = \kappa_v^{(\mathbb{Q})}(m_{t^-} - v_{t^-})dt + \sigma_v\sqrt{v_{t^-}}dW_t^{v(\mathbb{Q})} + dJ_t^{v(\mathbb{Q})} \quad (2)$$

$$dm_t = \kappa_m^{(\mathbb{Q})}(\theta_m^{(\mathbb{Q})} - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}dW_t^{m(\mathbb{Q})} + dJ_t^{m(\mathbb{Q})} \quad (3)$$

[18] In our model specification the market is not complete, therefore the risk-neutral measure is not unique. It will be estimated using market prices of underlying returns and options.

[19] Assuming that the interest rate r and dividend yield are constant, it does not matter which maturity of the forward we consider because the cash-and-carry relationship between the forward and the spot index ensures that all forwards have the same dynamics (but different initial conditions).

where W^Y, W^v, W^m are three $\mathbb{Q}$ Brownian motions and

$$d\langle W^Y, W^v\rangle_t = \rho_{Y,v}dt \ ; \ d\langle W^m, W^Y\rangle_t = 0 \ ; \ d\langle W^m, W^v\rangle_t = 0. \tag{4}$$

Our model is a two-factor stochastic volatility model with jumps, which allows the variance process $(v_t)_{t\geq 0}$ of the forward returns to revert towards a stochastic central tendency $(m_t)_{t\geq 0}$. The processes J^Y, J^v, J^m are finite activity jump processes defined by:

$$dJ_t^{Y(\mathbb{Q})} = Z_t^{Y(\mathbb{Q})}dN_t^{Yv} \ ; \ dJ_t^{v(\mathbb{Q})} = Z_t^{v(\mathbb{Q})}dN_t^{Yv} \ ; \ dJ_t^{m(\mathbb{Q})} = Z_t^{m(\mathbb{Q})}dN_t^{m}. \tag{5}$$

As suggested by the simultaneous peaks in the S&P 500 and VIX index, and in the expected forward returns on both indices, large movements in the equity returns and in the variance are likely to occur at the same time. Therefore we choose, in line with the literature (see, e.g., Eraker (2004), Broadie, Chernov, and Johannes (2007), Cont and Kokholm (2011)) to use the same Poisson process to generate jumps in the asset returns and in the variance process. We also choose the intensity of jumps to be dependent on the level of the factors. Formally, N_t^m and N_t^{Yv} are Poisson processes with respective intensities:

$$\lambda^m(m_{t^-}) = \lambda_0^m + \lambda_1^m m_{t^-} \tag{6}$$

$$\lambda^{Yv}(v_{t^-}, m_{t^-}) = \lambda_0^{Yv} + \lambda_1^{Yv} v_{t^-} + \lambda_2^{Yv} m_{t^-} \tag{7}$$

Moreover, the process $Z^{(\mathbb{Q})} = (Z^{Y(\mathbb{Q})}, Z^{v(\mathbb{Q})}, Z^{m(\mathbb{Q})})^\top$ corresponds to the random jump sizes under $\mathbb{Q}$ and we assume that their values taken at two times t and s are independent and identically distributed (i.i.d.) for all $t \neq s$. We assume that jump sizes in the forward returns are normally distributed $\mathcal{N}(\mu_Y^{(\mathbb{Q})}, \sigma_Y^{(\mathbb{Q})})$ and that the jump sizes in the two volatility factors are exponentially distributed with respective means $\nu_v^{(\mathbb{Q})}$ and $\nu_m^{(\mathbb{Q})}$. All jump sizes are independent from one another. These jumps sizes are characterized by their joint Laplace transform:

$$\theta_Z^{(\mathbb{Q})}(\phi) = \theta_Z^{(\mathbb{Q})}(\phi_Y, \phi_v, \phi_m) = \mathbb{E}^{\mathbb{Q}}[\exp(\phi^\top Z^{(\mathbb{Q})})], \tag{8}$$

where $\phi \in \mathbb{C}^3$.

This model implicitly defines dynamics for the VIX. In the following, we do not make the assumption that the VIX is approximately the 30-day realized volatility. Instead, we use its definition as a finite sum of call and put prices that converges (under the assumption that there exists call and put options

for all strikes in $\mathbb{R}_+$) to the integral

$$\begin{aligned}\text{VIX}_t^2 &= \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau}\frac{dF_u}{F_{u^-}} - d(\ln F_u)\right] \\ &= \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} v_u du + 2\left(e^{Z_u^{Y(\mathbb{Q})}} - 1 - Z_u^{Y(\mathbb{Q})}\right) dN_u^{Yv}\right],\end{aligned}$$

where τ is 30 days in annual terms.

In the affine model we use, the expression of the VIX is very simple and given below.

Proposition 4.1. *The VIX squared at time t can be written as an affine function of v_t and m_t:*

$$VIX_t^2 = \alpha_{VIX^2} v_t + \beta_{VIX^2} m_t + \gamma_{VIX^2} \tag{9}$$

where the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} are known in closed-form.

The coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} are provided in the Appendix 1.

4.2 Risk premium specification

We specify the change of measure from the pricing to the historical measure so that the model dynamics keep the same structure under $\mathbb{P}$. The parameters under $\mathbb{P}$ will simply have a superscript referring to the historical measure. We separate the total equity risk premium γ_t into a Brownian contribution which is proportional to the variance level and represents the compensation for the diffusive price risk, and a jump contribution which reflects the compensation for jump risk:

$$\gamma_t = \eta_Y v_{t^-} + \lambda^{Yv}(v_{t^-}, m_{t^-})\Big(\theta_Z^{(\mathbb{P})}(1,0,0) - \theta_Z^{(\mathbb{Q})}(1,0,0)\Big). \tag{10}$$

where $\theta_Z^{(\mathbb{P})}$ denotes the joint Laplace transform of jump sizes under the historical measure $\mathbb{P}$.

As in Pan (2002) and Eraker (2004) we impose the intensity of jumps to be the same under $\mathbb{Q}$ and $\mathbb{P}$.[20]

We define the mean price jump risk premium as the difference between the mean of the jump sizes in returns under $\mathbb{Q}$ and $\mathbb{P}$. Analogously, the volatility of price jump risk premium refers to the difference between the volatility of the jump sizes in returns under $\mathbb{Q}$ and $\mathbb{P}$.[21]

[20] Pan (2002) argues that introducing different intensities of jumps under the historical and pricing measure introduces a jump-timing risk premium that is very difficult to disentangle from the mean jump risk premium. The consequence of this assumption is that the jump-timing risk premium is artificially incorporated into the mean jump size risk premium.

[21] In the literature σ_Y has sometimes been constrained to be the same under $\mathbb{P}$ and $\mathbb{Q}$ (Bates (1988), Naik and Lee (1990)), but this is not required by absence of arbitrage and we follow Broadie, Chernov, and Johannes (2007) by allowing them to be different. Indeed, they find strong evidence for them to be different and report that this has strong implications for the magnitude of the premium attached to the mean price jump size.

We proceed similarly with the volatility risk premium and decompose it into a diffusive component and a jump component. The diffusive variance risk premium in v is proportional to the current level of variance, with coefficient of proportionality given by:

$$\eta_v = \kappa_v^{(\mathbb{Q})} - \kappa_v^{(\mathbb{P})}. \tag{11}$$

This risk premium should primarily be identified by the term structure of SPX implied volatilites as well as the cross section of VIX implied volatilities. The jump part of the volatility risk premium refers to the difference between the mean of the jump sizes in the variance under $\mathbb{Q}$ and $\mathbb{P}$.

Finally, we introduce a risk premium in the stochastic central tendency, which consists of a diffusive part proportional to the variance of m with coefficient of proportionality given by:

$$\eta_m = \kappa_m^{(\mathbb{Q})} - \kappa_m^{(\mathbb{P})}. \tag{12}$$

The corresponding jump risk premium in m is the difference between the mean of the jump sizes in m under $\mathbb{Q}$ and $\mathbb{P}$. The central tendency risk premia should be identified by the cross section and term structure of the VIX implied volatilities as well as the long-term SPX implied volatilities. Therefore, introducing options in our dataset with various moneynesses and maturities is crucial to have meaningful values for these risk premia.

Finally, no-arbitrage considerations force the volatility of volatilities (σ_v and σ_m) and the correlation between the returns and volatility $\rho_{Y,v}$ to be equal under $\mathbb{P}$ and $\mathbb{Q}$.

4.3 Derivatives pricing

Due to the affine property of the VIX^2 and given the results of Duffie, Pan, and Singleton (2000), we have the following result:

Proposition 4.2. *The Laplace transforms of the returns and VIX^2 defined by the model* (1) - (3) *are given by*

$$\begin{aligned}
\Psi_{VIX_T^2}(t, v_t, m_t; \omega) &:= \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega VIX_T^2}\right], \\
\Psi_{Y_T}(t, y_t, v_t, m_t; \omega) &:= \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega Y_T}\right],
\end{aligned}$$

are exponential affine in the factor processes:

$$\begin{aligned}
\Psi_{VIX_T^2}(t, v, m; \omega) &= e^{\alpha(T-t)+\beta(T-t)\cdot v+\gamma(T-t)\cdot m}, \\
\Psi_{Y_T}(t, y, v, m; \omega) &= e^{\alpha_Y(T-t)+\beta_Y(T-t)\cdot y+\gamma_Y(T-t)\cdot v+\delta_Y(T-t)\cdot m},
\end{aligned}$$

where α, β, γ, α_Y, β_Y, γ_Y and δ_Y are functions defined on $[0, T]$ by ODEs presented in the online Appendix 2. $\omega \in \mathbb{C}$ is chosen so that the expectations above are well defined.

In affine models, option pricing is most efficiently performed using Fourier inversion techniques since we know the Fourier transform of the stochastic processes of interest. To price options on the S&P 500, Fang and Oosterlee (2008) report that the Fourier Cosine Expansion is very efficient and fast compared to other Fourier inversion techniques. We use this method to price S&P 500 options and extend it to incorporate also the pricing of VIX options. This technique is comparable to the inversion performed by Sepp (2008a) but is more parsimonious in the number of computational parameters.

Pricing options on the VIX poses technical difficulties that are not encountered when pricing equity options. To understand why it is different, let us write the price of a call option with strike K and maturity T on the VIX at time $t = 0$ in the following form:

$$C(\text{VIX}_0, K, T) = e^{-rT} \int_0^\infty (\sqrt{v} - K)^+ f_{\text{VIX}_T^2}(v) dv, \tag{13}$$

where $f_{\text{VIX}_T^2}$ is the $\mathbb{Q}$ density of the VIX^2 at time $t = T$. We introduce the density of VIX^2 because this is the variable which is affine in our framework (as opposed to working with the VIX).

The square root appearing in the integral as part of the payoff prevents us from using the Fast Fourier Transform of Carr and Madan (1999). For S&P 500 call options the payoff can be written as $(e^y - K)^+$ where y is the log of the stock price. The fact that we have the exponential e^y makes it possible to interpret this integral as a Fourier transform. To apply the same methodology in the case of VIX derivatives, we would need the log of the VIX to be affine which is incompatible with affine models for log-returns. This justifies our choice to depart from the standard Fourier pricing techniques.

The basic idea of the method developed by Fang and Oosterlee (2008) is to write the density of the S&P 500 log-returns as a Fourier cosine expansion on a well chosen truncated interval $[a, b]$. This allows them to derive the price of S&P 500 options; we use the same methodology to calculate the price of VIX options[22].

Theorem 4.1. *Let us consider a European style contingent claim on the VIX index with maturity T and payoff $u_{VIX}(VIX^2) = (\sqrt{VIX^2} - K)^+$ (respectively on the normalized S&P 500 forward $\tilde{Y} := \log(F/K)$ with payoff $u_{SPX}(e^{\tilde{y}}) = K(e^{\tilde{y}} - 1)^+$ at T). Given a chosen interval $[a_{VIX}, b_{VIX}]$ for the support of the $VIX_T^2|_{v_0,m_0}$ density (respectively $[a_{\tilde{Y}}, b_{\tilde{Y}}]$ for the support of the density of $\tilde{Y}_T|_{Y_0} = \log(F_{t_0}(T)/K)|_{Y_0}$), the price $P_{VIX}(t_0, VIX_0)$ at time $t = t_0 \geq 0$ (respectively $P_{SPX}(t_0, Y_0)$) of the*

[22]For some parameter values, the cosine expansion for the density of the VIX^2 converges slowly. We found that this happens for parameters where the density of the VIX^2 is not differentiable at the left end of its support (close to γ_{VIX^2} in equation (9)) and this generates an oscillating Fourier approximation (referred to as the Gibbs phenomenon, well known in numerical analysis). One way to improve convergence is to use spectral filters as is illustrated in Ruijter, Versteegh, and Oosterlee (2013). Before one estimates the model, we cannot rule out the parameter values that generate the Gibbs phenomenon and therefore one needs to take this in consideration so that the optimizer can run over the whole space of parameters.

contingent claim is

$$P_{VIX}(t_0, VIX_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1}{}' A_n^{VIX^2} U_n^{VIX^2}, \tag{14}$$

$$\textit{respectively:} \quad P_{SPX}(t_0, Y_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1}{}' A_n^{\tilde{Y}} U_n^{\tilde{Y}}, \tag{15}$$

where the prime superscript in the sum $\sum'$ *means that the first term* A_0U_0 *is divided by 2. The terms in the sum are defined by:*

$$A_n^{VIX^2} = \frac{2}{b_{VIX} - a_{VIX}} Re\left\{ \Psi_{VIX_T^2}\left(t_0, v_0, m_0; \frac{in\pi}{b_{VIX} - a_{VIX}}\right) \exp\left(-ia_{VIX}\frac{n\pi}{b_{VIX} - a_{VIX}}\right)\right\}, \tag{16}$$

$$U_n^{VIX^2} = \int_{a_{VIX}}^{b_{VIX}} u_{VIX}(v) \cos\left(n\pi \frac{v - a_{VIX}}{b_{VIX} - a_{VIX}}\right) dv \tag{17}$$

respectively

$$A_n^{\tilde{Y}} = \frac{2}{b_{\tilde{Y}} - a_{\tilde{Y}}} Re\left\{ \Psi_{\tilde{Y}_T}\left(t_0, \tilde{Y}_0, v_0, m_0; \frac{in\pi}{b_{\tilde{Y}} - a_{\tilde{Y}}}\right) \exp\left(-ia_{\tilde{Y}}\frac{n\pi}{b_{\tilde{Y}} - a_{\tilde{Y}}}\right)\right\}, \tag{18}$$

$$U_n^{\tilde{Y}} = \int_{a_{\tilde{Y}}}^{b_{\tilde{Y}}} u_{SPX}(e^{\tilde{y}}) \cos\left(n\pi \frac{\tilde{y} - a_{\tilde{Y}}}{b_{\tilde{Y}} - a_{\tilde{Y}}}\right) d\tilde{y}, \tag{19}$$

where $\Psi_{\tilde{Y}_T}$ *is the Laplace transform of* $\tilde{Y}_T = Y_T - \log K$*, i.e.,*

$$\Psi_{\tilde{Y}_T}(t, \tilde{y}, v, m; \omega) = \Psi_{Y_T}(t, y, v, m; \omega) e^{-w \log K}.$$

We note that the coefficients $A_n^{\text{VIX}^2}$ and $A_n^{\tilde{Y}}$ are computed using Proposition 4.2 and the coefficient $U_n^{\text{VIX}^2}$ is known in closed form and given in Appendix 2. Finally, a closed form for $U_n^{\tilde{Y}}$ can be found in Fang and Oosterlee (2008).

5 Joint estimation and particle filter

The goal of this section is twofold. First, we explain how we calibrate the nested models (1) - (3) to S&P 500 and VIX options, i.e., we estimate the model under the pricing measure using the VIX and S&P 500 option price surfaces on a given date. This exercise allows us to show that the $\mathbb{Q}$ dynamics of the model is sufficiently rich to accurately price both S&P 500 and VIX derivatives together, i.e., at any date t we can find a fixed set of parameters which allows the model to price both VIX options and S&P 500 options accurately. Second and most importantly, we detail how we have built a time consistent estimation of the models using a time series of S&P 500 and VIX indices together with a time series of S&P 500 and VIX option prices. This means that we estimate both the $\mathbb{P}$ and $\mathbb{Q}$ dynamics

of the model using the time series of indices and options (i.e., we find one vector of parameters for the whole time series of SPX/VIX spots and SPX/VIX options). The algorithm we use is the Auxiliary Particle Filter, introduced by Pitt and Shephard (1999) which makes it possible for us to filter out unobserved latent variables, such as the volatility process or jumps.

From Section 4.1, we recall the $\mathbb{P}$- and $\mathbb{Q}$- parameter vectors:

$$\Theta^{\mathbb{P}} = \{\kappa_v^{(\mathbb{P})}, \kappa_m^{(\mathbb{P})}, \theta_m^{(\mathbb{P})}, \nu_m^{(\mathbb{P})}, \nu_v^{(\mathbb{P})}, \mu_Y^{(\mathbb{P})}, \sigma_Y^{(\mathbb{P})}, \eta_Y\} \tag{20}$$

$$\Theta^{\mathbb{Q}} = \{\kappa_v^{(\mathbb{Q})}, \kappa_m^{(\mathbb{Q})}, \theta_m^{(\mathbb{Q})}, \nu_m^{(\mathbb{Q})}, \nu_v^{(\mathbb{Q})}, \mu_Y^{(\mathbb{Q})}, \sigma_Y^{(\mathbb{Q})}\}. \tag{21}$$

The remaining parameters are equal under both measures:

$$\Theta^{\mathbb{P},\mathbb{Q}} = \{\lambda_0^{Yv}, \lambda_1^{Yv}, \lambda_2^{Yv}, \lambda_0^m, \lambda_1^m, \sigma_m, \sigma_v, \rho_{Yv}\}. \tag{22}$$

The vector of all parameters is then $\Theta = \{\Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}\}$.

5.1 Daily calibration - Methodology

In this approach, we calibrate our model to the cross section of S&P 500 and VIX options on some chosen dates. On each date, the output will be a set of values for the risk-neutral parameters $\Theta^{\mathbb{Q}}$ and $\Theta^{\mathbb{P},\mathbb{Q}}$. Calibration to one single day of options data does not allow us to estimate the parameters $\Theta^{\mathbb{P}}$ since options are priced under the pricing measure $\mathbb{Q}$ and no time series is used. This exercise is important because if the model is not able to reproduce accurately the implied volatility patterns of both markets together on a single date, then there is no point in estimating the model using a filter on a time series of options and indices.

We fix a date t. Let us consider $\{IV\_\text{SPX}_i^{Mkt}\}_{i=1\cdots I}$ the set of implied volatilities of options on the S&P 500 for the strikes $\{K_i\}$ and maturities $\{T_i\}$ available in our dataset[23] for this date. We use the superscript $_{Mkt}$ for 'Market' implied values. We denote by $\{IV\_\text{VIX}_j^{Mkt}\}_{j=1\cdots J}$ the set of VIX option implied volatilities on the same date t. I is the number of S&P 500 options available for this date and J the number of VIX options. To estimate parameter values, we minimize a distance between market and model implied volatilities (or option prices). We have chosen two distance criteria[24] that put different emphasis on S&P 500 and VIX options as well as on at-the-money (ATM) and out-of-the-money (OTM) options. We denote by $IV\_\text{SPX}_i^{Mod}$ the model implied volatility of option with strike K_i and maturity T_i (respectively $IV\_\text{VIX}_j^{Mod}$ corresponding to the notations above). The root

[23]The dataset is described in the empirical analysis section 3 where we explain how implied volatilities have been calculated from S&P 500 and VIX options.

[24]Since we analyze the fits in section 6 in terms of implied volatilities and not option prices, we do not consider other popular choices of distances including absolute error of the logarithm of option prices, relative error of option prices (see Christoffersen and Jacobs (2004)) . Alternatively, we checked that using distances taking into account the bid-ask spread of IVs as in Cont and Kokholm (2011) does not significantly change the quality of fits.

mean squared error (RMSE) in implied volatilities on date t is defined as:

$$\begin{aligned}
\text{RMSE_SPX}(t) &:= \sqrt{\frac{1}{I}\sum_{i\in I}\left(IV_\text{SPX}_i^{Mkt} - IV_\text{SPX}_i^{Mod}\right)^2} \\
\text{RMSE_VIX}(t) &:= \sqrt{\frac{1}{J}\sum_{j\in J}\left(IV_\text{VIX}_j^{Mkt} - IV_\text{VIX}_j^{Mod}\right)^2} \\
\text{RMSE}(t) &:= \frac{1}{2}\left(\text{RMSE_SPX}(t) + \text{RMSE_VIX}(t)\right)
\end{aligned} \tag{23}$$

We furthermore consider the average relative error (ARE) in implied volatilities on date t:

$$\begin{aligned}
\text{ARE_SPX}(t) &:= \frac{1}{\#I}\sum_{i\in I}\frac{\left|IV_\text{SPX}_i^{Mkt} - IV_\text{SPX}_i^{Mod}\right|}{IV_\text{SPX}_i^{Mkt}} \\
\text{ARE_VIX}(t) &:= \frac{1}{\#J}\sum_{j\in J}\frac{\left|IV_\text{VIX}_j^{Mkt} - IV_\text{VIX}_j^{Mod}\right|}{IV_\text{VIX}_j^{Mkt}} \\
\text{ARE}(t) &:= \frac{1}{2}\left(\text{ARE_SPX}(t) + \text{ARE_VIX}(t)\right).
\end{aligned} \tag{24}$$

Since the IVs are the highest for OTM SPX puts and OTM VIX calls, the RMSE puts more emphasis on fitting these options (which are the most liquid together with ATM options). On the other hand, it is also arguable that a 1% absolute error on an IV does not have the same importance if the market IV is 10% or 80%. The average relative error distance ARE takes this consideration into account by computing relative errors.

To cope with the ill-posedness of the calibration problem and the potential existence of multiple minima, we use two global optimizers namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996), and the Differential Evolution (DE) algorithm introduced by Storn (1996).[25] They are evolutionary algorithms designed for high-dimensional non-linear non-convex optimization problems in a continuous domain. They are based on the principle of biological evolution, i.e., at every step new vectors of parameters are generated based on the optimal set of parameters up to that step and random perturbations, the objective function is evaluated for each of these new parameter vectors, and the new optimal parameter set becomes the one which minimizes the objective function.

5.2 Particle filter

While the daily calibration provides us with a static estimation of parameter values, it is more insightful to use the whole time series of option prices and index levels to learn about the dynamic properties of the unobservable processes (volatility, central tendency, jumps) and the risk premia associated to them. Sequential Monte-Carlo methods are ideal for this purpose as they make it possible to progressively

[25] We are grateful to Jochen Krause for his implementation of various evolution optimizers including the CMA-ES and DE algorithms.

filter the trajectories of latent processes based on the information available. As they take as input a time series of observations, they furthermore allow to better identify parameters and therefore deliver more robust estimates. We discretize the continuous-time model under $\mathbb{P}$ on a uniform time grid composed of $M+1$ points $t \in \{t_0 = 0, t_1 = \Delta t, ..., t_k = k\Delta t, ..., t_M = M\Delta t\}$ (for some $M \in \mathbb{N}^*$). For any $t = t_k$, $(0 \leq k \leq M-1)$ we obtain the corresponding state-space discretization:

$$\begin{aligned}\Delta Y_t = Y_{t+\Delta t} - Y_t &= [-\lambda^{Yv}(v_t, m_t)(\theta_Z^{(\mathbb{P})}(1,0,0) - 1) - \frac{1}{2}v_t + \gamma_t]\Delta t + \sqrt{v_t}\Delta W_t^{Y(\mathbb{P})} \\ &\quad + Z_t^{Y(\mathbb{P})}\Delta N_t^{Yv} \qquad (25)\end{aligned}$$

$$\Delta v_t = v_{t+\Delta t} - v_t = \kappa_v^{(\mathbb{P})}\left(\frac{\kappa_v^{(\mathbb{Q})}}{\kappa_v^{(\mathbb{P})}}m_t - v_t\right)\Delta t + \sigma_v\sqrt{v_t}\Delta W_t^{v(\mathbb{P})} + Z_t^{v(\mathbb{P})}\Delta N_t^{Yv} \qquad (26)$$

$$\Delta m_t = m_{t+\Delta t} - m_t = \kappa_m^{(\mathbb{P})}(\theta_m^{(\mathbb{P})} - m_t)\Delta t + \sigma_m\sqrt{m_t}\Delta W_t^{m(\mathbb{P})} + Z_t^{m(\mathbb{P})}\Delta N_t^m. \qquad (27)$$

In practice, Δt will correspond to one day, since we use daily data. In particular, we do not augment the time space since Johannes, Polson, and Stroud (2009) have shown that the advantage of introducing additional time steps is very limited when using daily observations.

The latent factors we wish to infer from the observations are: $L_t = \{v_t, m_t, \Delta N_t^{Yv}, \Delta N_t^m, Z_t^{Y(\mathbb{P})}, Z_t^{v(\mathbb{P})}, Z_t^{m(\mathbb{P})}\}$. Note that among these factors, only v_t and m_t depend on their past values, as the jump sizes $\left(Z_t^{Y(\mathbb{P})}, Z_t^{v(\mathbb{P})}, Z_t^{m(\mathbb{P})}\right)$ are i.i.d. over time and so are the increments of Poisson processes conditionally on v_t and m_t. Equation (25) is the first measurement equation. The second one is given by the observation of the VIX index level with error. Indeed, since the VIX index is in practice calculated using a finite number of options, a discretization bias is introduced. Furthermore, Jiang and Tian (2007) point to systematic biases in the VIX. We write this error as follows:

$$\text{VIX}_t^2 - (\alpha_{\text{VIX}^2}v_t + \beta_{\text{VIX}^2}m_t + \gamma_{\text{VIX}^2}) = \epsilon_t^{\text{VIX}}. \qquad (28)$$

The error terms ϵ_t^{VIX} are assumed to follow a normal distribution with mean zero and variance $s > 0$. Other observable quantities are the prices of S&P 500 and VIX options. We assume that option prices are observed with an error, which is due to different sources such as the bid-ask spread, the processing and timing errors (all options considered on a given day are not traded at the same time) and misspecification error. The observation equations (29) - (30) for options are:

$$\frac{O_{t,i}^{\text{SPX},Mod}(Y_t, v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,i}^{\text{SPX},Mkt}}{O_{t,i}^{\text{SPX},Mkt}} = \epsilon_{t,i}^{\text{SPX},options} \qquad (29)$$

$$\frac{C_{t,j}^{\text{VIX},Mod}(v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}) - C_{t,j}^{\text{VIX},Mkt}}{C_{t,j}^{\text{VIX},Mkt}} = \epsilon_{t,j}^{\text{VIX},options} \qquad (30)$$

where $O_{t,i}^{\text{SPX},Mkt}$ corresponds to the market price at time t of the S&P 500 option indexed by $i \in I$, $O_{t,i}^{\text{SPX},Mod}(Y_t, v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}})$ to the model price of the same option assuming the $\mathbb{Q}$ parameters are $\{\Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}\}$. Similarly, $C_{t,j}^{\text{VIX},Mkt}$ denotes the market price at time t of the call option on the VIX

indexed by $j \in J$, and $C_{t,j}^{\text{VIX},Mod}(v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}})$ the model price with $\mathbb{Q}$ parameters $\{\Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}\}$. We assume the error terms to be normally distributed and heteroscedastic[26]:

$$\epsilon_{t,i}^{\text{SPX},options} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{t,i}^{\text{SPX}}}), \tag{31}$$

and

$$\epsilon_{t,j}^{\text{VIX},options} \sim \mathcal{N}(\mu_{\epsilon_t^{\text{VIX}}}, \sigma^2_{\epsilon_{t,j}^{\text{VIX}}}), \tag{32}$$

where $\mu_{\epsilon_t^{VIX}}$ is proportional to the error ϵ_t^{VIX} which has been made on the estimation of the VIX level. Indeed, if the underlying's value is not accurately estimated, it introduces a bias in the valuation of VIX options. We specify the variance of errors as follows:

$$\sigma^2_{\epsilon_{t,i}^{\text{SPX}}} = \exp\left(\phi_0 \text{bid-ask spread}_i + \phi_1 \left|\log\left(\frac{K_i}{F_t^{\text{SPX}}(T_i)}\right)\right| + \phi_2(T_i - t) + \phi_3\right) \tag{33}$$

$$\sigma^2_{\epsilon_{t,j}^{\text{VIX}}} = \exp\left(\psi_0 \text{bid-ask spread}_j + \psi_1 \left|\log\left(\frac{K_j}{F_t^{\text{VIX}}(T_j)}\right)\right| + \psi_2(T_j - t) + \psi_3\right). \tag{34}$$

with ϕ_i and ψ_i are in $\mathbb{R}$, $i \in \{0, ..., 3\}$.

At time $t = t_k$ (for $0 \leq k \leq M$), we denote by y_t, the set of observable prices. The log-likelihood of a time-series of $n+1$ observations $y^{t_n} = (y_{t_0}, ..., y_{t_n})$ $(n \leq M)$ with joint density p conditionally on a set of parameters Θ and a model specification $\mathcal{M}$ is equal to:

$$\log p(y^{t_n}|\Theta, \mathcal{M}) = \log p(y_{t_0}, ..., y_{t_n}|\Theta, \mathcal{M}) = \sum_{k=1}^{n} \log p(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) + \log p(y_{t_0}|\Theta, \mathcal{M}) \tag{35}$$

where, by the Law of Total Probability,

$$p(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) = \int p(y_{t_k}|L_{t_k}, \Theta, \mathcal{M}) p(L_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) dL_{t_k}. \tag{36}$$

Given an initial density $p(L_{t_0}|\Theta, \mathcal{M})$, the transition density of state variables $p(L_{t_k}|L_{t_{k-1}}, \Theta, \mathcal{M})$ and the likelihood function $p(y_{t_k}|L_{t_k}, \Theta, \mathcal{M})$, filtering methods make it possible to estimate the distribution $p(L_{t_k}|y^{t_k}, \Theta, \mathcal{M})$ of the current state at time t_k given all observations up to that time. In particular, particle filters are perfectly adapted to our problem since they can handle observations which are nonlinear functions of latent variables as well as non-Gaussian innovations. The filtering density is given by Bayes' formula as follows:

$$p(L_{t_k}|y^{t_k}) \propto p(y_{t_k}|L_{t_k}) p(L_{t_k}|y^{t_{k-1}}). \tag{37}$$

[26]The fact that option pricing errors are normally distributed does not constitute a restriction. The reason is that the errors are heteroscedastic and coefficients generating heteroskedasticity are driven by the data, i.e., we optimize over the parameters $\{\phi_i, \psi_i\}_{0 \leq i \leq 3}$.

The likelihood function is known, but the predictive distribution of the state is not. It is given by the following integral, which involves the previous filtering density:

$$p(L_{t_k}|y^{t_{k-1}}) = \int p(L_{t_k}|L_{t_{k-1}})p(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}}. \tag{38}$$

The key idea is to approximate the posterior density function of the latent variables $p(L_{t_k}|y^{t_k}, \Theta)$ by a sum of point masses positioned at strategic points called particles $\{L_{t_k}^{(i)}\}_{1\leq i\leq n_p}$:

$$\hat{p}(L_{t_k}|y^{t_k}, \Theta, \mathcal{M}) = \sum_{i=1}^{n_p} \pi_{t_k}^{(i)} \delta(L_{t_k} - L_{t_k}^{(i)}) \tag{39}$$

where $\pi_{t_k}^{(i)}$ denotes the normalized importance weight for particle i and $\delta(.)$ is the Dirac function. n_p is the number of support points (particles) for $\hat{p}(L_{t_k}|y^{t_k}, \Theta, \mathcal{M})$. Then the filtering density is recursively calculated as follows:

$$\begin{aligned}\hat{p}(L_{t_k}|y^{t_k}) &\propto \int p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}})\hat{p}(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}} \\ &= \sum_{i=1}^{n_p} p(y^{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}}^{(i)})\pi_{t_{k-1}}^{(i)}.\end{aligned} \tag{40}$$

To apply the particle filter, ones needs to be able to simulate at every time t_k a number n_p of particles $L_{t_k}^{(i)}, i = 1, ..., n_p$ from $p(L_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M})$ and to be able to evaluate $p(y_{t_k}|L_{t_k}^{(i)}, \Theta, \mathcal{M})$. Based on these simulated particles, $p(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M})$ is approximated by:

$$p(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) \approx \frac{1}{n_p} \sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k}^{(i)}, \Theta, \mathcal{M}). \tag{41}$$

Multiple versions of the particle filter exist. We use the Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (1999) and extend the approach described in Johannes, Polson, and Stroud (2009). The main advantage of the APF compared to more basic particle filters such as the Sampling Importance Resampling (SIR) filter is that it is more capable of detecting jumps compared to the SIR filter which faces sample impoverishment leading to particle degeneracy. Both filters are described in Johannes, Polson, and Stroud (2009) for filtering latent factors from returns in a Heston model with jumps. Our extension of the filter makes it possible to extract the most probable paths of both factors v and m as well as the jump components from the set of observable variables y^{t_n}.

To incorporate the information contained in the S&P 500 and VIX levels, we calculate at every time t the joint probability of having 0 or 1 jump in every process given the new observations. Since the jump size of returns is normally distributed but jumps in the variance processes are exponentially distributed, the conditional likelihood of the new observations given a combination of jumps involves the sum of

a normal and (up to two) exponential random jumps. To compute the joint probability of jumps and preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (generalization of a Bernoulli distribution) which has support a certain number of chosen quantiles.[27]

The detailed filtering procedure is described in Appendix 6.

6 Daily calibration results

The first step in evaluating the performance of the model (1)-(3) is to calibrate it to one day of options data to make sure that the model is flexible enough to simultaneously price options on both markets.

We follow the method outlined in section 5.1. We have chosen some dates on which we calibrate the model (1)-(3) to the cross section of S&P 500 and VIX implied volatilities. We only report the results on four dates as they are representative of the whole sample. We consider two days where the market was facing great uncertainty about the future, October 22, 2008 (one month after the bankruptcy of Lehman Brothers) and May 05, 2010, at the beginning of the European sovereign debt crisis. On these dates, the markets were under stress and S&P 500 implied volatilities had a very strong negative skew and levels above 100% for short term options. The other two days we report are rather calm compared to those: July 11, 2007 and June 10, 2009.

Gatheral (2008) has shown that the Heston model is incapable of reproducing the positive skew in VIX implied volatilities (IVs) as displayed in Figure 4 and Sepp (2008a,b) added that introducing positive jumps in the volatility dynamics of the Heston model allows the model to have a positive skew in VIX IVs. As a consequence, departing from the usual literature on S&P 500 option pricing, the simplest model we consider is the Heston model with jumps in returns and volatility. It corresponds to the model (1)-(2) where the central tendency m is constant. We will denote this model by SVJ (Stochastic Volatility with Jumps). The most flexible model we consider, with 2 factors to represent the volatility component namely the variance and a stochastic central tendency, is referred to as SVJ2.

[Insert Figure 4 here]

We report in Table 3 the results for the RMSE calibration with respect to implied volatilities (23).[28] We emphasize that for each model and on each day, we have minimized the total RMSE from the VIX and the S&P 500 market together. We report the resulting $RMSE_$SPX and $RMSE_$VIX since these are indicative of the quality of the fit on each market. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting very well with an average RMSE of around 1.5% across the dates we have calibrated to. In contrast, we see that there are dates when the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs

[27] Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.
[28] The results are qualitatively similar when minimizing over the distances (23) and (24), we therefore only report one result.

whereas the SVJ2 model satisfactorily fits both, for instance on July 11, 2007 and on May 05, 2010. We can see the comparative fits for the VIX options market in Figure 5.[29] Panels A to D correspond to the SVJ2 model's fit and panels E to H to the SVJ model's fit. On this date, it seems that for short maturities, the SVJ model has difficulties reproducing the strong positive skewness of VIX IVs (which was already the case for the Heston model). This shortcoming of the SVJ is not often noticeable and we therefore do not make it a general statement. Indeed, on the other two dates we report, the fits of the SVJ and SVJ2 are comparable on the VIX market.

[Insert Table 3 here]

[Insert Figure 5 here]

Daily calibration is essentially a multiple curve fitting exercise, where we check whether the models can fit the risk-neutral distributions inferred by option prices at different maturity. A more thorough analysis is needed to conclude at this point that the SVJ2 is much better than the SVJ model to price S&P 500 and VIX options together. Indeed the SVJ2 has more parameters and is therefore bound to fit better. Furthermore, some of the parameters we get from daily calibrations can vary a lot from one day to the next.[30] At this point, it is therefore not possible to know whether the dynamics of the model (1)-(3) can reproduce the time evolution of these smiles. This is what we will focus in the next section.

7 Time-series estimation results using a Particle Filter

The second step in evaluating the performance of the model is to estimate it using the time series of S&P 500 and VIX indices together with S&P 500 and VIX options. This is achieved using the particle filter described in Section 5.2. We report the results for different sub-models of (1)-(3) and analyze the gain of information and robustness we have from adding the data from the VIX market to the dataset. The sub-models considered of the SVJ2 model are the SVJ model and the 2-factor continuous model, which has a stochastic central tendency but no jumps. We refer to the latter model as SV2. As the likelihood results and pricing errors were not significantly different with and without jumps in the stochastic central tendency, all results reported for the SVJ2 model are without jumps in the m process.

We detail in Appendix 4 the specific data treatment performed before running the particle filter, as well as a description of the option prices used in the in- and out-of-sample periods. We divide the resulting data into four different datasets:

Dataset 1: S&P 500 returns and VIX index levels,

[29] The SVJ and SVJ2 both match the S&P 500 options market prices almost perfectly so we do not show them.
[30] As explained in Broadie, Chernov, and Johannes (2007) and Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008), the parameters obtained when calibrating to daily options prices are not stable over time. To better understand how the model performs over time, it is important to estimate the model on a time series of options data.

Dataset 2: S&P 500 returns, VIX index levels and S&P 500 options,

Dataset 3: S&P 500 returns, VIX index levels and VIX options,

Dataset 4: S&P 500 returns, VIX index levels, S&P 500 options and VIX options.

Using these different datasets allows us to make inference on what information they contain, and whether they are consistent with one another.

7.1 Filtered trajectories

Figure 7 displays the filtered trajectory of the volatility process when estimating all three models SVJ, SV2 and SVJ2 over *Dataset 4*. Panel D of Figure 9 shows the jump sizes when the filtered jump probability is larger than 50%. We note that the volatility trajectories are consistent across models but that the SVJ generates more volatility jumps during the crisis so that the volatility can stay high without a long term volatility factor. Indeed, the Feller condition[31] is imposed on the SVJ model, which restricts the amplitude of volatility movements. This condition is relaxed for 2-factor models as the long-term mean of the variance is varying, which allows the volatility process to have a larger amplitude. In particular, for all datasets the estimated parameter σ_v, which controls the volatility of volatility, is considerably smaller for the SVJ model than for the SV2 and SVJ2 models, see Table 5. Figure 8 represents the difference between the filtered volatility processes using *Dataset 4* and the other datasets (*Datasets 1* to *3*). Until the peak in the VIX toward the end of the in-sample period, this difference is very small (lower than 2%). In this period, the filtered volatility using *Dataset 1* remains slightly lower than the volatility filtered using the other datasets so options seem to contribute to increase the filtered volatility only moderately. When the VIX index reaches its highest peak at the end of 2008, the volatility filtered using *Dataset 4* is on average close to the one filtered using *Dataset 2* but up to 25% lower than the volatility filtered using *Datasets 1* and *3*. It appears that adding S&P 500 options to *Dataset 1* provides significant new information on the behavior of the variance, which is not spanned by VIX options. In the out-of-sample period, the difference between the trajectories remains within an interval of $\pm 3\%$ except during the second smaller peak of variance in May 2010. As for the contribution of jumps to the variation of the variance, Figure 9 shows that many jumps are filtered during the crisis regardless of the dataset used for estimation. Small jumps (around 2%) are filtered in the variance process at the beginning of 2007, and larger jumps (above 10%) are filtered towards the end of 2008 as the VIX peaks.

[Insert Figures 7, 8, 9 here]

[Insert Table 5 here]

The process m is overall more stable and less erratic than the variance process, giving evidence that it captures long-term trends. In fact, its volatility parameter σ_m is in the range $[10\%, 25\%]$ for

[31] We impose the Feller condition $2\kappa_v^{\mathbb{Q}} m_0 \geq \sigma_v^2$ on the SVJ model, where m_0 is the level of reversion of the variance when the central tendency is a constant process.

both the SV2 and SVJ2 models regardless of the dataset chosen for the estimation. Its speed of mean-reversion is more than ten times smaller than that of v under $\mathbb{P}$ and can be more than fifteen times smaller under $\mathbb{Q}$. The process is therefore more persistent. While the variance process increases dramatically during the crisis but then returns to a level which is comparable to the one before the crisis, the central tendency also increases during the crisis but then goes down to a level which is higher than before the crisis (between 5% and 8% after the crisis, compared to 3% before). Therefore v captures the punctual movements of the variance while m, being the stochastic long-term reversion level of v, embeds longer-term expectations of investors regarding the variance and can therefore be seen as a smoothed indicator of market turmoil. Even though the process m is more stable than v, the approximation made by the SVJ model of the process m being a constant is too rough. The level which the central tendency reaches during and after the crisis is clearly underestimated by the SVJ model. In fact, the constant central tendency estimated in the SVJ model seems to be close to the average filtered central tendency of the SV2 and SVJ2 models over the in-sample period. This makes the SVJ model insensitive and non-adaptable to different regimes in long-term volatility especially when the out-of-sample estimation period exhibits more instabilities than the in-sample period.

[Insert Figure 10 here]

7.2 Are jumps and/or a stochastic central tendency needed?

The filtered trajectories of the central tendency m for models SV2 and SVJ2 show that this process is clearly not constant. However, it could be the case that having m stochastic has a limited impact in terms of pricing and forecasting performances. In this section, we precisely analyze whether jumps and a stochastic central tendency are needed to reproduce the features of VIX levels, S&P 500 option prices and finally VIX option prices.

Let us first investigate whether these features are needed to provide an accurate fit of the VIX index. Given the filtered trajectories for the processes v and m inferred by *Dataset 1* we calculate the corresponding model-implied values of the VIX index using the optimal parameters. As illustrated by Figure 11, the model accurately reproduces the time-series of VIX index values. Table 6 reports the corresponding Mean Errors (ME) and Root Mean Square Errors (RMSE) and shows that they are comparable across models. This observation is consistent across datasets. At first sight, jumps and a stochastic central tendency therefore seem superfluous to reproduce the trajectory of the VIX level.

[Insert Figure 11 here]

[Insert Table 6 here]

To statistically challenge this claim, we run likelihood tests. Table 5 reports the log-likelihood values as well as the values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC)

for *Datasets 1* to *4*. When focusing on *Dataset 1*, both criteria are slightly in favor of the SVJ2 model and therefore support the use of jumps and of a stochastic central tendency.

We further challenge the in- and out-of-sample performance of the SVJ2 model by running various Diebold-Mariano (DM) tests. For the three models, we consider two loss functions namely the absolute and quadratic pricing errors, respectively defined as $L(e_t) = |e_t|$ and $L(e_t) = e_t^2$, where e_t refers to the difference at time t between the model-forecast of the VIX and the true value of the index. We denote the loss differential at time t between the SVJ or the SV2 model and the SVJ2 model by $d^{SVJ/SV2,SVJ2} = L(e_t^{SVJ/SV2}) - L(e_t^{SVJ2})$. If the two models considered have comparable pricing errors, then $\mathbb{E}^{\mathbb{P}}[d^{SVJ/SV2,SVJ2}] = 0$. A positive value of the expectation means that the SVJ2 model outperforms the sub-model considered. Table 7 reports the results when the calibration is done using *Dataset 1* and shows that the test values are very close to zero with the quadratic loss function and negative with the absolute loss function, suggesting that the SVJ and SV2 models should be preferred to the SVJ2 model when calibrating them to underlying levels only. When including options in the estimation dataset, still evaluating only the fit of the VIX index time series, we obtain test values which are very close to zero (within ± 0.1 bounds). Therefore we conclude that the SVJ2 model does not significantly outperform the SVJ2 and SV2 models at reproducing the trajectories of VIX levels, in- and out-of-sample.

[Insert Table 7 here]

The calibration to S&P 500 options (*Dataset 2*) highlights the superiority of the SVJ2 and SV2 models over the SVJ model. The SVJ model exhibits Root Mean Square Errors (RMSEs) and Root Mean Square Relative Errors (RMSREs) - when pricing S&P 500 options - which are for most option categories higher than those of the SV2 and SVJ2 models. In particular, the SVJ model does not represent well the deep OTM and long-maturity options, see Tables 4 and 8. The corresponding RMSREs are about one and a half those of the 2-factor volatility models for in- and out-of-sample deep OTM calls and long-maturity options. Therefore a stochastic central tendency allows to better price long-term and deep OTM S&P 500 options. This supports the hypothesis that the process m captures the long-term trends of volatility, and therefore makes it possible to better reproduce the term structure of S&P 500 option prices.

[Insert Table 8 here]

The calibration to VIX options (*Dataset 3*) also favors the SVJ2 model, which yields better RMSEs and RMSREs of VIX option prices than the SVJ and SV2 models except for out-of-sample deep OTM and short-maturity options. The comparison of the SV2 and SVJ models shows that the SVJ model better prices these deep OTM calls while the SV2 model is more appropriate for other moneyness levels. The SVJ model can therefore better represent the tail of the volatility distribution. Consistently with the results we obtained when estimating models to *Dataset 2*, the SV2 model outperforms the SVJ model in pricing the medium-maturity VIX options (which are the longest maturity VIX options).

When including options on both markets in the estimation dataset (*Dataset 4*), the SVJ2 model yields RMSEs and RMSREs which are smaller than the SVJ and the SV2 models in-sample in pricing most S&P 500 and VIX option categories, see Tables 4 and 8. We notice that the SVJ model performs particularly worse at pricing the deep OTM options on the S&P 500, consistently with the results obtained when calibrating the models to *Dataset 2*.

Using the Diebold-Mariano test, we investigate whether the SVJ2 model's pricing performance of SPX and VIX options is significantly better than that of its sub-models on average. For this purpose, we consider two loss functions, the Mean Square Error (MSE) of option prices and the Mean Square Relative Error (MSRE). The resulting test values are presented in Tables 10 and 9. They confirm that the SVJ2 model provides significantly better in-sample MSREs for S&P 500 options than the other two models. The test values associated to VIX option pricing are essentially positive. We however note that in the out-of-sample period, the test indicates that the SVJ2 model does not outperform the SV2 model, which might be due to a problem of identification of the jump terms, as a large part of the crisis belongs to the out-of-sample time period.

[Insert Tables 9 and 10 here]

We conclude that a stochastic central tendency adds significant value for pricing long-term options and the representation the tails of the returns' distribution (OTM calls on the S&P 500). On the other side, jumps add value to represent the right tail of the variance distribution (OTM calls on the VIX) as well as short-term options. Therefore the jumps and the stochastic central tendency of the SVJ2 model provide important improvements over the SVJ and SV2 to represent the underlying returns in a way that is consistent with S&P 500 returns, VIX levels and their derivatives prices. This conclusion is however mitigated by the difficulty to identify jump terms.

Furthermore we observe that the SVJ2 model encounters problems during the crisis and does not well represent volatility smiles. In particular, OTM puts on the S&P 500 tend to be underpriced and OTM calls are generally overpriced, i.e., the smile of volatility does not exhibit enough skewness. This phenomenon affects short-maturity options in particular. Figure 12 compares the moments of S&P 500 returns as implied by market and model option prices, when the models are calibrated to *Dataset 4* (all indices and options). While the skewness of the returns is well represented at the beginning of the in-sample period, it is underestimated from late 2007 until the end of our sample. In the out-of-sample period this phenomenon becomes much more apparent, and all three models yield an implied skewness which is about half the one implied by the market. The SVJ2 model provides a slight improvement over the other two models but is still far from reality. Similarly, the kurtosis is only slightly underestimated at the beginning of the time-series, but in the out-of-sample period the model kurtosis it is about half the market implied kurtosis. We add that there is no improvement in the representation of SPX implied moments when adding the options on the VIX market to the estimation dataset.

[Insert Figure 12 here]

7.3 Market integration

In this part we address the following two questions: first, what information do levels contain on options and second, which information do S&P 500 and VIX options share?

Since the value of the VIX index is calculated using a portfolio of S&P 500 options, it is tempting to see the VIX index value as a summary of the information contained in S&P 500 options. In fact, Figure 8 shows that the volatilities filtered are the same across estimation datasets before the crisis. This seems to contradict previous results (e.g., Johannes, Polson, and Stroud (2009)) stating that including options in the estimation dataset increases the estimated volatility level. We attribute this difference to the fact that our *Dataset 1* is not only composed of returns (as in Johannes, Polson, and Stroud (2009)) but also contains VIX index levels, which incorporates some information about the options market. In fact, our results suggest that in the calm periods, *Dataset 1* is sufficient to recover the trajectory of the volatility. Following such reasoning, it is interesting to see to which extent the model estimated using *Dataset 1* is able to reproduce options' prices. We obtain RMSEs and RMSREs that are significantly larger than those obtained using *Dataset 4* for both S&P 500 and VIX options (even in the calmer first period). Our results indicate that even though estimating the model to returns and VIX index time series seems to be sufficient to filter volatility in "good times", it is definitely not sufficient to estimate appropriately model parameters. This leads to strong mispricing of options in both the SPX and VIX markets, even in the in-sample period. We conclude that estimating the model to underlying index values is not sufficient to reproduce option prices in either market and to extract the dynamics of the variance process process accurately in times of market stress.

[Insert Figure 13 here]

Although we have seen that the VIX and S&P 500 options provide conflicting information on the trajectory of volatility in times of market turmoil, the jumps filtered in the SVJ2 model present the same patterns. A small jump (around 2%) is filtered at the beginning of 2007, then towards the end of the in-sample period, in fall 2008, a couple of jumps between 10 and 25% occur with probability larger than 0.5. Finally, a jump of about 13% is detected when the Eurozone sovereign debt crisis emerged in May 2010. Despite this, we note that the RMSEs and RMSREs on S&P 500 options are overall much lower (reduced by a factor 2 to 3 approximately) estimating to *Dataset 2* compared to estimating to *Dataset 3*. This is in particular due to the fact that estimating to VIX options leads to strongly mispriced deep OTM SPX calls, which indicates that VIX options contain less information on the right tail of the returns' distribution than S&P 500 options. Concerning deep OTM puts on the S&P 500, it is striking to see that the estimation using *Dataset 3* outperforms the one using *Dataset 2* out-of-sample, which indicates that VIX options provide valuable information on the left tail of the returns' distribution.

Conversely, we note that the RMSEs and RMSREs of VIX options using *Dataset 2* are about one and a half those using *Dataset 3*, see Tables 4 and 8. This consideration holds in- and out-of-sample. Thus we conclude that S&P 500 options do not span the information contained in VIX options.

There is a significant loss of quality in the fitting of VIX options when estimating the SVJ2 model to all data sources, which indicates that even the SVJ2 model is not flexible enough to reconcile the information contained in all data sources in a completely consistent way. This contrasts with the results obtained in the static calibrations performed in section 6.

Our results therefore suggest that on the one side there is a loss of information when summarizing S&P 500 option prices in the VIX index, and that on the other side the S&P 500 and VIX option markets are not integrated. This result has two major consequences. The first is the potential mispricing of options on one market if one estimates the model using the other markets' data. The second consequence of this lack of integration is related to hedging. If for instance a trader wants to use VIX options to hedge the volatility in a portfolio of SPX options, it is important to have a model sufficiently flexible to match prices on both markets in a dynamic way, otherwise the Greeks will be completely wrong.

7.4 Risk premia

We first analyze the signs of the risk premia defined in Section 4.2. The equity risk premium coefficient η_Y is found to be positive across models and datasets considered, which is in line with a positive diffusive equity risk premium. When options are part of the estimation dataset, it is consistently between 0.6 and 0.85. The mean price jump risk premium is slightly negative, i.e., the mean jump size of returns is slightly more negative under $\mathbb{Q}$ (around -10%) than under $\mathbb{P}$ (around -4%), which is also what Pan (2002) finds. We do not find evidence of a volatility of price jump risk premium being significantly away from zero, the volatility of jump sizes under both measures is around 10%. The diffusive part of the volatility risk premium is found to be negative, its amplitude however depends on the model used. In particular, it has much smaller magnitude with the SVJ2 and SV2 models than with the SVJ model. Indeed, for the first two models, the diffusive part of the stochastic central tendency risk premium adds to the volatility risk premium, and is also found to be negative. Finally, when we include options in the estimation dataset, we find that the mean volatility jump risk premium is in general negative, as the average jump size in the variance is between 17 and 20% under $\mathbb{P}$ and between 8 and 14% under $\mathbb{Q}$.

We now examine integrated risk premia and their term structure as implied by the different models, calibrated over the four datasets previously considered.

Consistently with Bollerslev and Todorov (2011) and Aït-Sahalia, Karaman, and Mancini (2012), the annualized integrated equity risk premium (IERP) is defined as follows:

$$IERP_t = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T \frac{dF_s}{F_s}\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T \frac{dF_s}{F_s}\right)\right].$$

Similarly, the annualized integrated variance risk premium (IVRP) is defined as:

$$IVRP(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(QV_{[t,T]}\right) - \mathbb{E}_t^{\mathbb{Q}}\left(QV_{[t,T]}\right)\right]$$

where $QV_{[t,T]}$ denotes the quadratic variation of the log price process, which is the sum of the integrated variance of the returns and the squared jumps in the time interval considered:

$$QV_{[t,T]} = \int_t^T v_s ds + \sum_{t \leq s \leq T} J_s^2 \Delta N_s^{Yv}.$$

The IERP represents the expected excess return on the market for an investor who holds S&P 500 Futures from t to T. The IVRP represents the payoff that one would expect when buying a variance swap at time t with maturity T. Typically, the IERP is found to be positive, i.e. the investor requires to be compensated for the risk in future fluctuations of the stock. In contrast, the IVRP is usually found negative, which can be explained by looking at the variance swap as an insurance product against future fluctuations in volatility, which therefore embeds an insurance premium.

The IERP and IVRP can be decomposed into a continuous and a discontinuous part, the first one generated by the diffusion of the log-returns and the second by the jump component. As the jump intensity is an affine function of the latent factors v and m, all calculations boil down to a linear combination of $\int_t^T \mathbb{E}_t^{\mathbb{P}}[v_{s^-}]ds$ and $\int_t^T \mathbb{E}_t^{\mathbb{P}}[m_{s^-}]ds$, which are known in closed form. As a consequence, the continuous and jump components of both risk premia at time t can be decomposed as a linear combination of v_t and m_t. The continuous and discontinuous parts of both risk premia as well as the contribution of the two factors are given in the Online Appendix 7.

We obtain integrated equity risk premia that are positive and strongly increasing during the crisis in 2008, then coming back to a low level and suddenly increasing again end of 2010, following the VIX movements, see Figure 14. During the periods of market turmoil, shorter-maturity risk premia are larger than longer-maturity premia which is in line with Aït-Sahalia, Karaman, and Mancini (2012). Indeed, at the peak of a crisis investors perceive a short-term investment as being more risky than a longer investment, which has larger probability to end after the market turmoil is over. When the economy is doing well, it is the other way around and agents find it safer to invest over a shorter period of time. Although the IERP is positive across datasets and models, we find that it has larger magnitude when S&P 500 options are part of the estimation dataset. When options are not present, the value of η_Y is smaller, and being proportional to the continuous IERP, it scales down the equity premium. Our explanation is that S&P 500 options contain better information than VIX options on the S&P 500 returns' distribution and therefore on the IERP. Besides, we find that the equity risk premium is primarily determined by its continuous part, for all maturities. Taking as example the instantaneous IERP, the jump part is smaller than 0.1 for all datasets considered and all models while the continuous part reaches around 0.7. Finally, we find that the stochastic central tendency has negligible impact on the continuous component of the IERP for all maturities considered. In contrast, the central tendency process is main driver of the (small) discontinuous part of the IERP.

Regarding the IVRP, they are negative across estimation datasets and models, consistently with the literature. The magnitude increases in 2008, and we find that shorter-maturity risk premia are closer to zero than longer-maturity premia, as shown on Panel A of Figure 15. Furthermore, our results suggest that the discontinuous component of the IVRP dominates for shorter maturities (see Panel B of Figure 15), while the continuous component becomes more important for longer term maturities, which is an additional argument in favour of jumps in the dynamics of the underlying to accurately represent the shorter end of the variance term structure. Indeed, as jumps in the variance are constrained to be positive, a jump in the process will make the payoff of a variance swap jump as well, with small chances to get back to the previous level if the maturity is close. With a longer maturity, the process is likely to revert to its long-term mean and therefore the IVRP is less sensitive to the occurrence of a jump. Finally, we find that the contribution of the stochastic central tendency in the continuous part of the IVRP is small for close maturities but no longer negligible for maturities larger than three months. In quiet times, the contribution of m sets the level of the continuous IVRP, while the contribution of v takes over when the variance peaks, see Panel C of Figure 15. m also has a non-negligible impact on the discontinuous part of the IVRP and determines a large part of its level in quiet times, see Panel D of Figure 15. During the crisis, the peak of the discontinuous IVRP is also driven by the shorter-term variance factor v.

[Insert Figures 14 and 15 here]

We conclude that jumps and the stochastic central tendency have a minor impact on the IERP but are important to estimate the IVRP. While jumps help represent the shorter end of the variance term structure, the stochastic central tendency m plays a determinant role in setting the level of the IVRP when the market is calm.

8 Conclusion

In this paper we estimate a general affine model with jumps using a time series of S&P 500 and VIX levels as well as option prices on both indices. To extract as much information about extreme events as possible, we use S&P 500 and VIX options with a unique wide range of moneynesses. We depart from most of the literature and estimate the historical $\mathbb{P}$-parameters and the risk-neutral $\mathbb{Q}$-parameters jointly, in a single step. This estimation puts less restrictions on the parameters and allows us to obtain results on risk premia which are more data-driven. We show that although the standard SVJ model performs well at representing the smiles of volatility for both markets on a given date, its dynamics is not sufficiently flexible to accommodate for the dynamical properties embedded in the time series of option prices. We argue that a model with a stochastic central tendency and jumps in the returns and in the variance factor brings significant improvements and makes it possible to reach smaller pricing errors, both in and out-of-sample. We analyze the filtered trajectories of the latent factors using different estimation datasets and sub-models. We show that the variance process exhibits large and fast variations capturing the short-term movements of the volatility while the stochastic central

tendency exhibits more persistence and hence reflects long-term expectations of investors. We provide an extensive analysis of which features of the SVJ2 are needed to represent different datasets. In particular, likelihood criteria as well as statistical tests conclude that the flexibility of the model is needed to jointly represent the index levels and the derivatives' prices on both markets. Indeed, adding a stochastic central tendency helps to better represent the tails of the returns' distribution as well as the term structure of S&P 500 and VIX option prices. Furthermore, jumps make it possible to better reflect the right tail of the variance distribution as well as short-dated options. However, based on likelihood criteria and pricing errors, we do not find evidence that jumps in the central tendency factor add value to capture the time series of option prices. We highlight the limitations of the models considered, in particular we show that they are not able to fully reproduce the skewness and kurtosis of the underlying S&P 500 index in times of market turmoil. We investigate the information contained in the underlying levels and in the options on both markets. On the one hand, we find that the VIX index does not provide an accurate representation of the information contained in S&P 500 options, and on the other hand that the information contained in S&P 500 derivatives does not span the information contained in VIX derivatives and vice-versa. It is therefore crucial to include underlyings as well as derivatives on both markets to estimate a model and account for the cross section of instruments. We finally provide a discussion on the risk premia embedded in the model. We find that all the datasets considered provide consistent information on the equity risk premium and that it is mainly determined by its continuous component. We emphasize the importance of jumps and of a stochastic central tendency in representing the term structure and level of variance risk premia.

Appendix

1 Affine dependence of the VIX^2 on v and m

We provide the expression of the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} in Proposition 4.1. We can write

$$\begin{aligned}
\alpha_{\text{VIX}^2} &= \left(1 + 2\lambda_1^{Yv} C\right) A \\
\beta_{\text{VIX}^2} &= \left(1 + 2\lambda_1^{Yv} C\right) B + \left(2\lambda_2^{Yv} C\right) \hat{A} \\
\gamma_{\text{VIX}^2} &= 2\lambda_0^{Yv} C + \left(1 + 2\lambda_1^{Yv} C\right) G + \left(2\lambda_2^{Yv} C\right) \hat{B}.
\end{aligned}$$

where $C := \left(\theta_Z^{(\mathbb{Q})}(1,0,0) - \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_Y}(0,0,0) - 1\right)$ and coefficient $A, B, \hat{A}, \hat{B}, G$ are defined in Table 11 of the online Appendix. Furthermore, $A, B, \hat{A}, \hat{B}, G$ are functions of:

$$\begin{aligned}
a_m &:= \left(\frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_m}(0,0,0)\lambda_1^m - \kappa_m^{(\mathbb{Q})}\right), \\
c_m &:= \left(\kappa_m^{(\mathbb{Q})}\theta_m^{(\mathbb{Q})} + \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_m}(0,0,0)\lambda_0^m\right), \\
b_m &:= -\frac{c_m}{a_m}\ , \quad \text{when } a_m \neq 0, \\
a_v &:= \left(\frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0)\lambda_1^{Yv} - \kappa_v^{(\mathbb{Q})}\right), \\
b_v &:= b_m\left(\kappa_v^{(\mathbb{Q})} + \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0)\lambda_2^{Yv}\right) + \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0)\lambda_0^{Yv}\ , \quad \text{when } a_m \neq 0, \\
h_v &:= \left(\kappa_v^{(\mathbb{Q})} + \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0)\lambda_2^{Yv}\right).
\end{aligned}$$

2 Coefficients for the Fourier Cosine Expansion

Here we give the expression for $U_n^{\text{VIX}^2}$, the Fourier cosine transform of VIX options' payoff. To ease notation, we drop the subscript VIX for $a_{\text{VIX}}, b_{\text{VIX}}$ and define $\omega_n := \frac{n\pi}{b-a}$. $U_n^{\text{VIX}^2}$ is given by:

$$\begin{aligned}
U_n^{\text{VIX}^2} &= \int_a^b \left(\sqrt{x} - K\right)^+ cos\left(\omega_n(x-a)\right) dx \\
&= \frac{2}{b-a} Re\left\{e^{-i\omega_n a}\left[\sqrt{b}\frac{e^{-i\omega_n b}}{i\omega_n} + \frac{\sqrt{\pi}}{2}\frac{1}{(-i\omega_n)^{3/2}}\left(\text{erfz}(\sqrt{-i\omega_n b}) - \text{erfz}(K\sqrt{-i\omega_n})\right)\right]\right\} \quad \text{for } n > 0,
\end{aligned} \tag{42}$$

$$U_0^{\text{VIX}^2} = \frac{2}{b-a}\left[\frac{2}{3}b^{3/2} - Kb + \frac{1}{3}K^3\right]. \tag{43}$$

3 Tables and Figures

3.1 Tables

Table 1: Descriptive Statistics for S&P 500 Futures log-returns and VIX levels

	March 2006 - November 2008				December 2008 - October 2010			
	Mean	Standard dev.	Skewness	Kurtosis	Mean	Standard dev.	Skewness	Kurtosis
Log-returns SPX	-0.0007	0.0159	0.0722	14.0772	0.0007	0.0158	-0.3283	6.8683
VIX	0.2044	0.1208	2.6560	10.9620	0.2907	0.1025	1.1929	3.8631

Table 2: Quantiles of implied volatility values for S&P 500 and VIX options. The quantiles are calculated using all maturities and moneyness available from March 2006 to October 2010.

Quantiles	SPX IVs	VIX IVs
25%	16.93%	62.95%
50%	23.16%	75.03%
75%	32.77%	91.68%

Table 3: Root Mean Squared Error in implied volatilities when calibrating the SVJ and the SVJ2 models to S&P 500 and VIX options on four different dates. On each date, the distance (23) is minimized using a global optimizer over the model parameters under $\mathbb{Q}$. On each date, the data is composed of all S&P 500 options implied volatilities and all VIX implied volatilities together. Here we report the distances $RMSE\_SPX$ and $RMSE\_VIX$ but the minimization is run over $RMSE = \frac{1}{2}(RMSE\_SPX + RMSE\_VIX)$.

RMSE (%)	20070711		20081022		20090610		20100505	
	SPX	VIX	SPX	VIX	SPX	VIX	SPX	VIX
SVJ	1.766	8.007	2.508	13.601	1.870	9.997	1.271	11.599
SVJ2	0.852	3.885	2.427	12.757	2.110	7.933	1.169	5.153

Table 4: Root Mean Square Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated to, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column. The moneyness $Mon.$ is defined here as the ratio between the strike of the option and the corresponding maturity futures' price. TTM denotes the time-to-maturity of options.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4			
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	#
Fitting SPX options in the in-sample period													
Overall	11.873	6.999	9.128	3.253	6.359	3.524	4.727	4.410	6.281	2.912	5.375	2.840	14351
$Mon. < 0.7$	2.902	4.423	3.680	3.146	3.816	3.376	2.224	2.026	2.715	2.791	3.365	2.746	1365
$0.7 \leq Mon. \leq 0.95$	6.286	7.203	5.283	3.478	6.501	3.875	2.401	3.350	2.458	2.766	5.484	2.878	6498
$0.95 \leq Mon. \leq 1.05$	13.604	7.731	10.880	3.173	6.946	3.184	5.177	4.501	5.874	3.327	5.809	3.015	3673
$1.05 \leq Mon. \leq 1.2$	19.631	6.845	14.456	2.803	6.536	2.954	7.589	5.415	11.217	2.745	5.494	2.502	2287
$Mon. > 1.2$	19.154	4.894	12.559	3.005	4.554	3.864	9.372	10.579	13.923	2.567	4.516	2.720	528
$TTM < 2$ months	2.619	2.073	2.457	1.359	1.729	1.268	2.282	2.692	2.753	1.405	1.673	1.367	4476
2 months $\leq TTM < 6$ months	7.856	5.384	6.428	2.545	4.820	2.600	3.787	3.691	5.263	2.366	3.982	2.371	5499
$TTM > 6$ months	19.436	10.970	14.669	4.967	10.018	5.530	7.067	6.264	9.317	4.332	8.483	4.179	4376
Fitting SPX options in the out-of-sample period													
Overall	9.974	11.030	9.330	5.141	10.045	5.663	7.411	6.122	9.203	4.390	8.739	4.228	13323
$Mon. < 0.7$	3.639	5.892	4.778	4.039	5.349	4.214	2.367	3.855	2.981	3.612	4.991	3.511	3046
$0.7 \leq Mon. \leq 0.95$	6.167	13.221	8.121	6.259	11.586	6.789	4.263	6.921	4.373	5.242	10.168	5.089	4755
$0.95 \leq Mon. \leq 1.05$	11.373	13.220	10.730	5.073	12.061	5.783	8.899	7.242	10.300	4.331	10.195	4.179	1962
$1.05 \leq Mon. \leq 1.2$	15.317	10.725	12.873	4.408	10.588	5.117	11.448	6.374	14.604	3.820	9.029	3.553	2349
$Mon. > 1.2$	15.723	7.729	11.548	4.059	7.608	4.709	11.676	4.810	15.770	3.547	6.709	3.341	1211
$TTM < 2$ months	2.732	2.219	2.609	1.518	1.833	1.611	2.823	3.614	3.442	1.543	1.950	1.668	3580
2 months $\leq TTM < 6$ months	7.158	7.195	6.753	3.275	6.099	3.551	5.444	3.766	7.106	2.783	5.094	2.789	5678
$TTM > 6$ months	15.745	18.015	14.684	8.343	16.606	9.230	11.472	9.567	14.021	7.088	14.515	6.728	4065
Fitting VIX options in the in-sample period													
Overall	0.021	0.026	0.022	0.018	0.022	0.021	0.012	0.014	0.013	0.015	0.019	0.015	4148
$Mon. > 1.3$	0.008	0.007	0.007	0.006	0.007	0.007	0.005	0.004	0.005	0.006	0.007	0.006	1100
$1.1 \leq Mon. \leq 1.3$	0.014	0.013	0.012	0.010	0.012	0.011	0.007	0.007	0.007	0.009	0.012	0.009	963
$0.9 \leq Mon. \leq 1.1$	0.021	0.020	0.018	0.014	0.019	0.016	0.008	0.011	0.009	0.012	0.017	0.012	1075
$Mon. < 0.9$	0.032	0.046	0.039	0.031	0.038	0.037	0.021	0.025	0.023	0.026	0.032	0.025	1010
$TTM < 2$ months	0.017	0.022	0.020	0.016	0.019	0.017	0.011	0.014	0.012	0.014	0.017	0.013	1948
$TTM \geq 2$ months	0.023	0.029	0.024	0.020	0.025	0.024	0.013	0.014	0.013	0.017	0.021	0.016	2200
Fitting VIX options in the out-of-sample period													
Overall	0.025	0.039	0.032	0.028	0.037	0.028	0.017	0.024	0.017	0.025	0.035	0.024	5118
$Mon. > 1.3$	0.007	0.009	0.009	0.009	0.009	0.009	0.007	0.007	0.007	0.009	0.009	0.009	2039
$1.1 \leq Mon. \leq 1.3$	0.014	0.020	0.018	0.017	0.020	0.017	0.011	0.015	0.011	0.016	0.020	0.016	916
$0.9 \leq Mon. \leq 1.1$	0.025	0.035	0.029	0.026	0.035	0.026	0.016	0.025	0.016	0.024	0.034	0.023	1109
$Mon. < 0.9$	0.046	0.076	0.059	0.051	0.071	0.052	0.030	0.044	0.032	0.045	0.065	0.043	1054
$TTM < 2$ months	0.022	0.032	0.028	0.024	0.030	0.024	0.014	0.017	0.014	0.021	0.028	0.021	1054
$TTM \geq 2$ months	0.027	0.044	0.034	0.030	0.042	0.031	0.019	0.029	0.019	0.027	0.039	0.026	2237

Table 5: Estimated parameters from the particle filter for the different models and datasets.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
						$\mathbb{P}$ and $\mathbb{Q}$ parameters						
λ_0^{Yv}	0.394	0.030	0	0.067	0.072	0	0.150	0.099	0	0.082	0.088	0
λ_1^{Yv}	1.650	0.007	0	0.497	2.225	0	0.368	2.386	0	0.246	2.376	0
λ_2^{Yv}	2.438	0	0	2.460	0	0	2.016	0	0	2.558	0	0
σ_m	0.150	0	0.147	0.095	0	0.131	0.248	0	0.209	0.120	0	0.208
σ_v	0.715	0.718	0.717	0.763	0.394	0.918	0.704	0.340	0.765	0.763	0.375	0.776
ρ_{Yv}	-0.550	-0.625	-0.578	-0.836	-0.946	-0.786	-0.779	-0.812	-0.446	-0.862	-0.934	-0.822
m_0		0.027			0.021			0.021			0.023	
						$\mathbb{P}$ parameters						
κ_v^P	10.002	11.057	9.723	7.768	8.162	7.259	7.790	7.502	8.231	7.378	7.558	7.622
κ_m^P	0.060	0	0.057	0.607	0	0.505	0.496	0	0.452	0.819	0	0.932
θ_m^P	0.501	0	0.504	0.009	0	0.021	0.077	0	0.049	0.011	0	0.027
ν_v^P	0.112	0.155	0	0.169	0.209	0	0.182	0.178	0	0.170	0.204	0
μ_Y^P	-0.060	0.150	0	-0.020	-0.080	0	-0.060	-0.061	0	-0.020	-0.040	0
σ_Y^P	0.041	0.007	0	0.061	0.043	0	0.100	0.042	0	0.110	0.034	0
η_Y	0.721	0.000	0.733	0.710	0.761	0.507	0.300	0.563	0.450	0.730	0.782	0.592
						$\mathbb{Q}$ parameters						
κ_v^Q	10.002	9.799	10.379	6.619	5.361	7.835	6.747	3.659	7.403	6.048	4.423	5.728
κ_m^Q	0.111	0	0.108	0.287	0	0.486	0.406	0	0.169	0.326	0	0.402
θ_m^Q	0.272	0	0.266	0.019	0	0.022	0.095	0	0.130	0.026	0	0.064
ν_v^Q	0.157	0.763	0	0.089	0.144	0	0.142	0.250	0	0.092	0.128	0
μ_Y^Q	0.040	-0.120	0	-0.070	-0.101	0	-0.090	-0.116	0	-0.070	-0.118	0
σ_Y^Q	0.033	0.006	0	0.112	0.139	0	0.110	0.090	0	0.109	0.114	0
Log-likelihood	10626	10572	10520	9692	4901	8357	9918	8386	9123	6700	356	3855
AIC	-21206	-21116	-21020	-19342	-9768	-16688	-19790	-16734	-18216	-13346	-668	-7674
BIC	-21103	-21053	-20975	-19248	-9692	-16630	-19687	-16649	-18149	-13225	-569	-7593

Table 6: Statistics on in-sample and out-of-sample errors on VIX levels, for the different models and datasets.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
In-sample fitting levels												
Mean error VIX	-0.001	-0.001	-0.001	0.002	0.003	0.001	0.000	-0.001	0.001	0.002	0.002	0.002
RMSE VIX	0.023	0.023	0.022	0.021	0.020	0.021	0.023	0.020	0.023	0.021	0.021	0.021
Out-of-sample fitting levels												
Mean error VIX	0.001	0.001	0.001	0.003	0.003	0.002	-0.001	-0.004	-0.001	0.002	0.002	0.002
RMSE VIX	0.022	0.022	0.021	0.023	0.022	0.023	0.021	0.022	0.021	0.023	0.022	0.023

Table 7: Diebold-Mariano test values for in-sample and out-sample errors on VIX levels, for different models calibrated to log-returns and VIX levels. Two loss functions are considered: the absolute and quadratic error. Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	SVJ		SV2	
	Abs loss function	Quad loss function	Abs loss function	Quad loss function
In-sample	-0.278	-0.001	-0.253	-0.001
Out-of-sample	-0.243	-0.002	-0.253	-0.002

Table 8: Root Mean Square Relative Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated on, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column. The moneyness $Mon.$ is defined here as the ratio between the strike of the option and the corresponding maturity futures' price. TTM denotes the time-to-maturity of options.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4			
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	#
Fitting SPX options in the in-sample period													
Overall	2.946	0.850	1.849	0.250	0.364	0.266	0.798	0.853	1.561	0.259	0.370	0.298	14351
$Mon. < 0.7$	0.738	0.625	0.573	0.404	0.570	0.413	0.521	0.898	0.522	0.390	0.578	0.407	1365
$0.7 \leq Mon. \leq 0.95$	0.501	0.398	0.449	0.220	0.326	0.265	0.296	0.353	0.414	0.204	0.318	0.272	6498
$0.95 \leq Mon. \leq 1.05$	0.675	0.367	0.484	0.191	0.200	0.171	0.262	0.347	0.343	0.230	0.217	0.216	3673
$1.05 \leq Mon. \leq 1.2$	5.221	1.779	3.246	0.250	0.395	0.260	1.161	1.660	2.676	0.286	0.400	0.335	2287
$Mon. > 1.2$	10.500	1.396	6.505	0.390	0.713	0.362	3.034	1.840	5.617	0.435	0.764	0.512	528
$TTM <$ 2 months	1.785	0.979	1.151	0.296	0.355	0.339	0.707	1.150	1.425	0.311	0.383	0.391	4476
2 months $\leq TTM <$ 6 months	2.798	0.797	1.725	0.227	0.331	0.218	0.655	0.675	1.527	0.231	0.323	0.238	5499
$TTM >$ 6 months	3.922	0.767	2.473	0.223	0.411	0.235	1.018	0.678	1.727	0.230	0.409	0.252	4376
Fitting SPX options in the out-of-sample period													
Overall	2.337	0.751	1.631	0.427	0.587	0.450	1.368	0.990	2.219	0.404	0.573	0.448	13323
$Mon. < 0.7$	0.729	0.898	0.812	0.706	0.833	0.713	0.537	0.647	0.716	0.656	0.806	0.673	3046
$0.7 \leq Mon. \leq 0.95$	0.455	0.590	0.482	0.355	0.484	0.393	0.247	0.336	0.431	0.324	0.446	0.387	4755
$0.95 \leq Mon. \leq 1.05$	0.307	0.379	0.298	0.150	0.287	0.171	0.238	0.490	0.297	0.152	0.297	0.184	1962
$1.05 \leq Mon. \leq 1.2$	2.744	1.006	1.988	0.232	0.513	0.281	1.491	1.710	2.560	0.247	0.523	0.297	2349
$Mon. > 1.2$	6.571	0.772	4.344	0.335	0.693	0.377	3.899	1.795	6.270	0.374	0.727	0.507	1211
$TTM <$ 2 months	1.427	0.992	1.096	0.462	0.565	0.510	0.990	1.514	1.595	0.457	0.605	0.537	3580
2 months $\leq TTM <$ 6 months	2.447	0.609	1.633	0.407	0.549	0.412	1.375	0.823	2.405	0.377	0.522	0.400	5678
$TTM >$ 6 months	2.783	0.674	1.982	0.422	0.654	0.444	1.620	0.494	2.413	0.389	0.611	0.425	4065
Fitting VIX options in the in-sample period													
Overall	1.088	0.738	0.795	0.495	0.639	0.596	0.266	0.375	0.310	0.416	0.586	0.435	4148
$Mon. > 1.3$	1.309	0.764	0.835	0.637	0.724	0.708	0.414	0.447	0.495	0.558	0.682	0.576	1100
$1.1 \leq Mon. \leq 1.3$	1.218	0.811	0.898	0.516	0.672	0.663	0.210	0.382	0.226	0.418	0.625	0.456	963
$0.9 \leq Mon. \leq 1.1$	1.054	0.751	0.867	0.404	0.595	0.526	0.158	0.334	0.188	0.318	0.537	0.344	1075
$Mon. < 0.9$	0.629	0.609	0.514	0.369	0.548	0.449	0.188	0.318	0.203	0.313	0.473	0.303	1010
$TTM <$ 2 months	1.101	0.761	0.909	0.492	0.617	0.612	0.297	0.352	0.342	0.409	0.579	0.462	1948
$TTM \geq$ 2 months	1.075	0.717	0.677	0.497	0.658	0.582	0.235	0.394	0.279	0.422	0.591	0.410	2200
Fitting VIX options in the out-of-sample period													
Overall	0.650	0.894	0.793	0.748	0.896	0.742	0.545	0.611	0.605	0.699	0.868	0.697	5118
$Mon. > 1.3$	0.714	0.916	0.946	0.914	0.936	0.905	0.757	0.619	0.853	0.885	0.925	0.894	2039
$1.1 \leq Mon. \leq 1.3$	0.652	0.915	0.756	0.734	0.912	0.718	0.413	0.659	0.435	0.664	0.891	0.651	916
$0.9 \leq Mon. \leq 1.1$	0.646	0.873	0.675	0.596	0.878	0.594	0.312	0.615	0.327	0.521	0.838	0.501	1109
$Mon. < 0.9$	0.505	0.855	0.588	0.509	0.821	0.519	0.286	0.544	0.307	0.439	0.760	0.422	1054
$TTM <$ 2 months	0.666	0.846	0.798	0.720	0.854	0.711	0.542	0.534	0.619	0.675	0.824	0.684	1054
$TTM \geq$ 2 months	0.637	0.930	0.789	0.769	0.927	0.766	0.547	0.665	0.595	0.718	0.901	0.707	2237

Table 9: Diebold-Mariano test values for in-sample and out-sample errors on VIX option prices, for the different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE). Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	Dataset 3				Dataset 4			
	SVJ		SV2		SVJ		SV2	
	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE
In-sample	0.000	0.110	0.000	0.360	0.000	0.354	0.000	0.382
Out-of-sample	0.000	0.021	0.000	0.535	0.000	0.451	0.000	-0.409

Table 10: Diebold-Mariano test values for in-sample and out-sample errors on S&P 500 option prices, for different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE) . Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	Dataset 2				Dataset 4			
	SVJ		SV2		SVJ		SV2	
	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE
In-sample	65.437	0.130	9.986	0.172	57.423	0.118	86.592	0.200
Out-of-sample	59.570	0.310	13.333	0.246	49.302	0.315	-5.166	0.134

3.2 Figures

Figure 1: Joint evolution of the VIX (solid curve) and SPX index (dashed curve) values from February 2006 to June 2010. The left y-axis corresponds to the VIX values (in percentage) and the right y-axis to the S&P 500 index values.

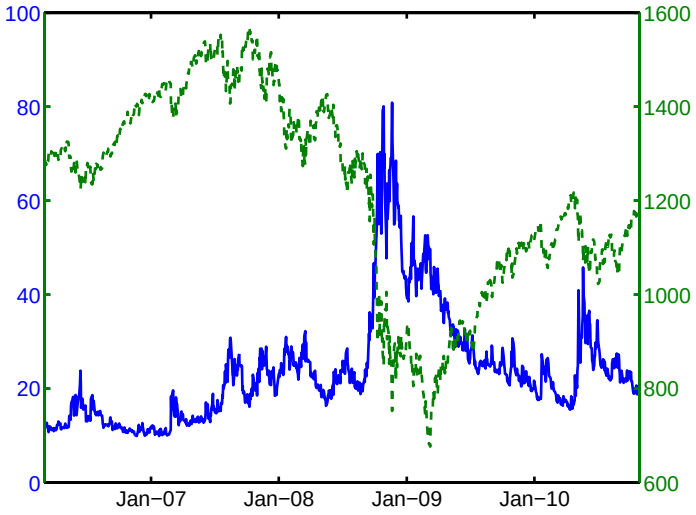


Figure 2: SPX and VIX implied volatility skews on four different dates as a function of log-moneyness ($\log K/F$). For each market, the scale is the same across days. The maturities T are quoted in years.

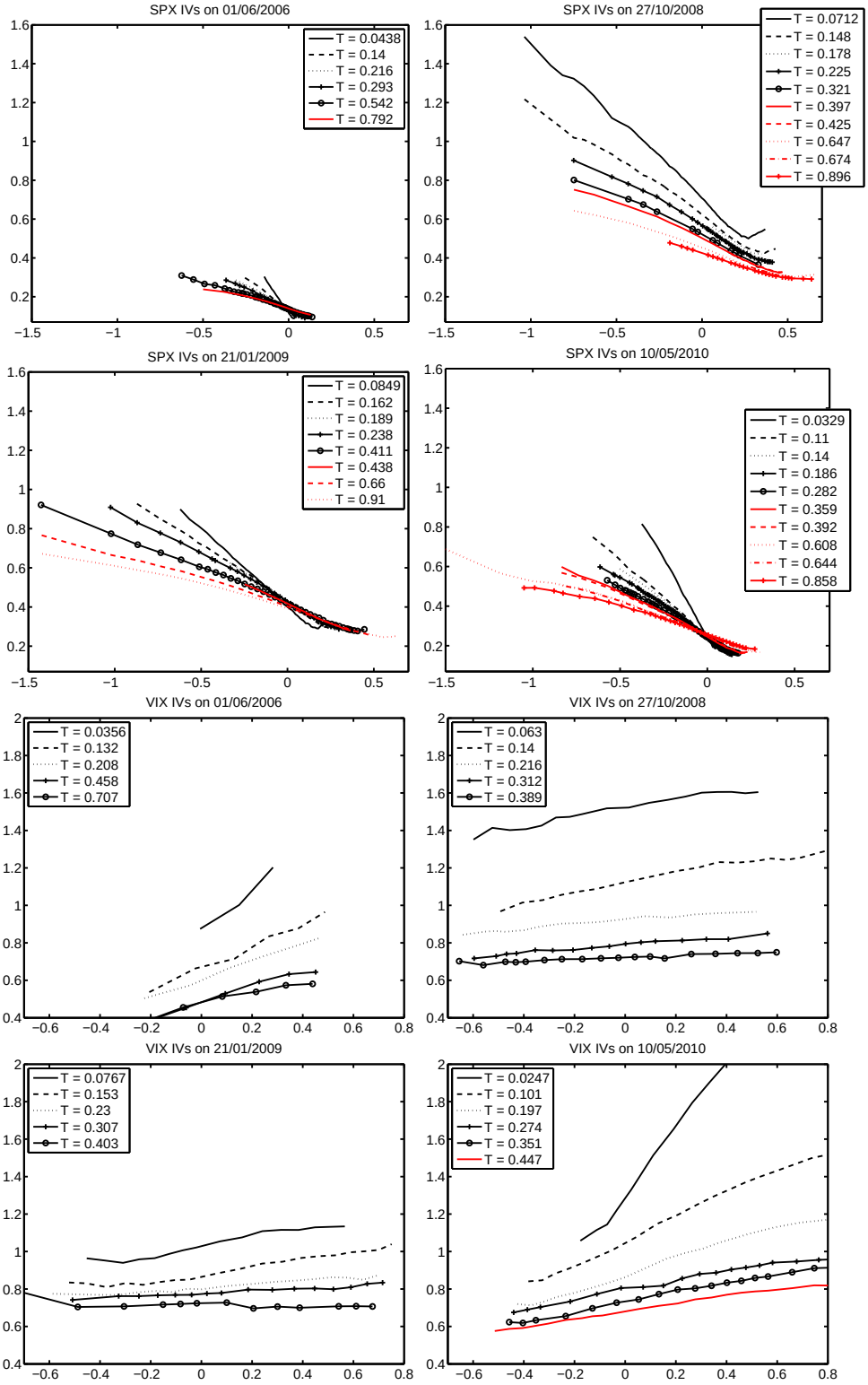


Figure 3: One month expected returns of the S&P 500 Futures implied by S&P 500 options with maturity one month from March 1st, 2006 to October 29th, 2010. We use the method described in Bakshi, Kapadia, and Madan (2003). Returns are expressed in percentage units.

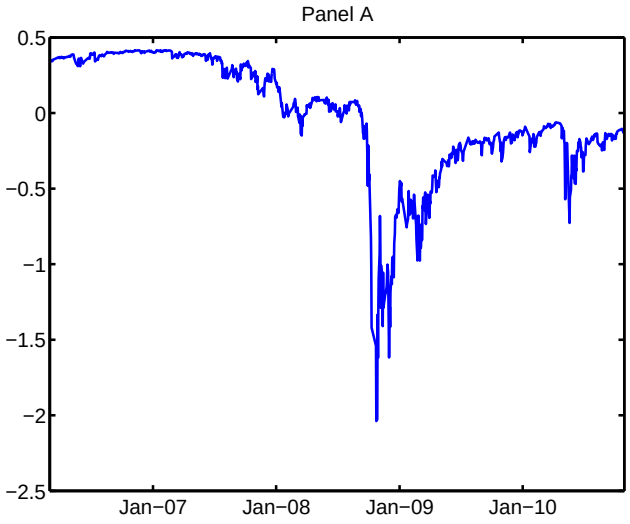


Figure 4: Market and Heston model implied volatilities for VIX options (four maturities) on October 20th, 2010 plotted with respect to forward log-moneyness ($\log K/F(T)$). The market (resp. model) implied volatilities are represented by the crosses (resp. the solid line). These fits are obtained by minimizing relative errors between market implied volatilities and the Heston model implied volatility.

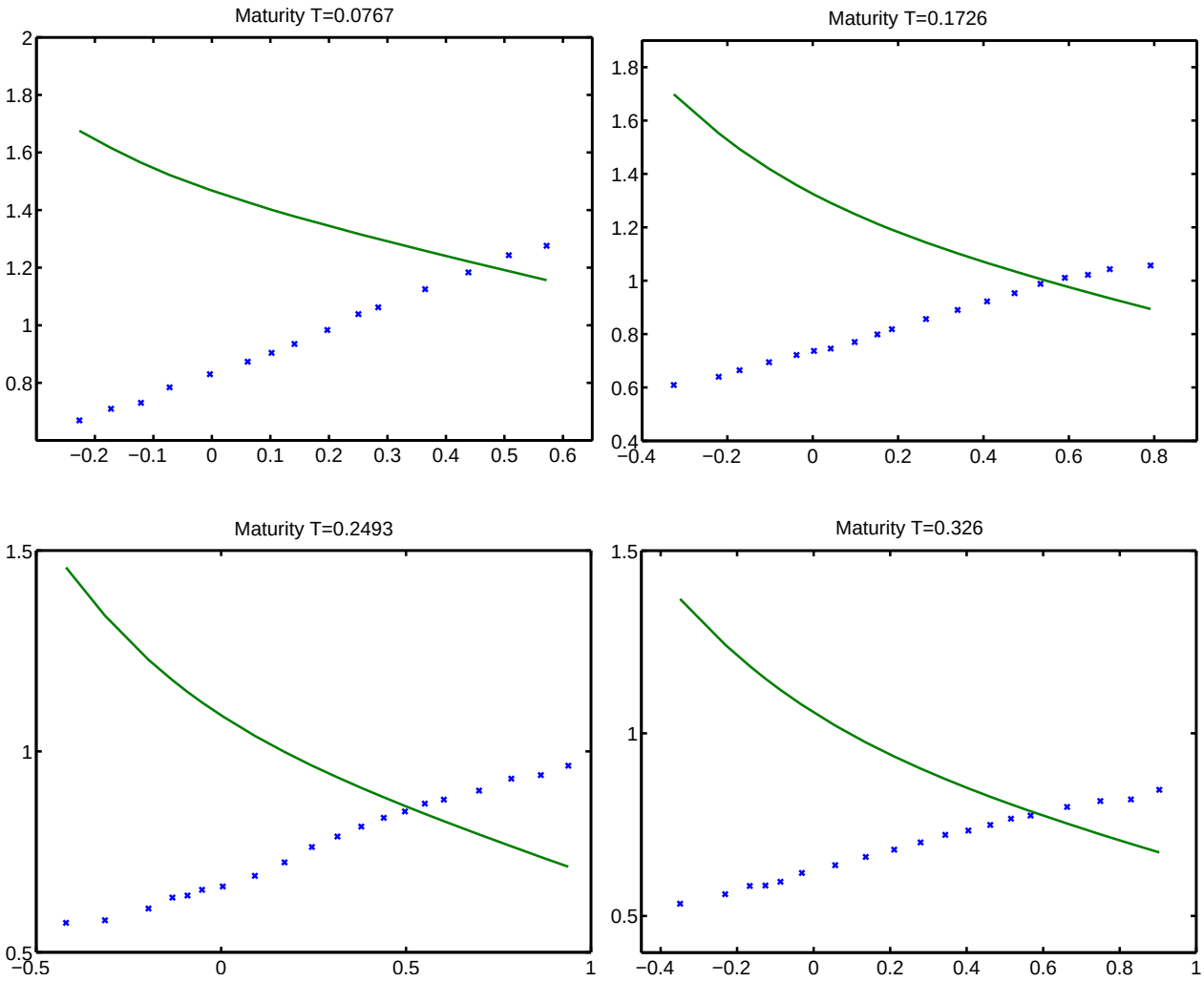


Figure 5: Comparative fit of 4 different maturities of VIX options for the SVJ2 (panels A to D) and the SVJ models (panels E to H) on July 11th, 2007. The crosses are the market implied volatilities and the curve represents the model volatilities. The implied volatilities are plotted as a function of forward log-moneyness ($\log K/F(T)$).

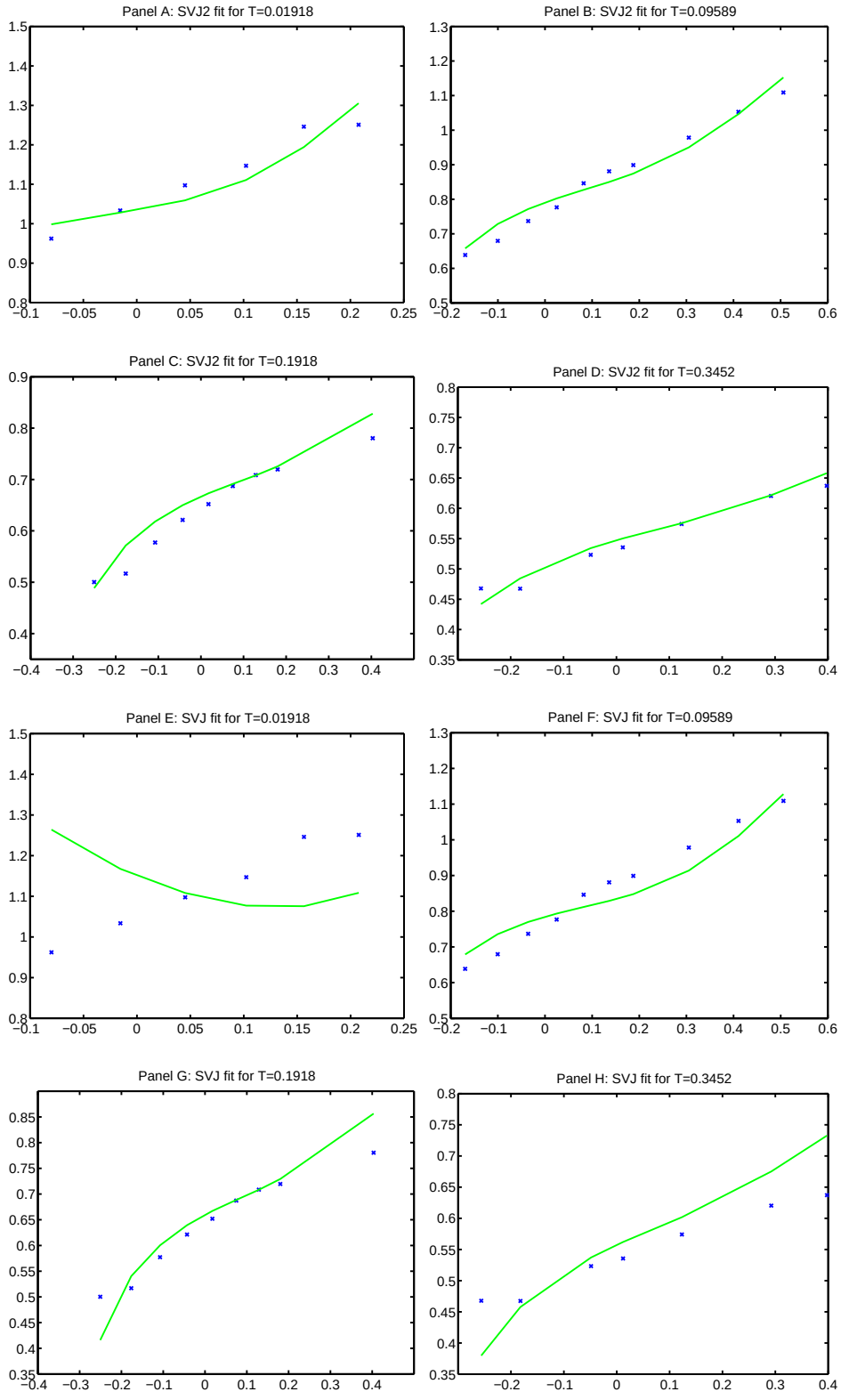


Figure 6: Typical interpolation of market implied volatilities (circles) using a mixture of log-normal densities for the density of Futures prices. The implied volatilities are plotted as a function of forward log-moneyness ($\log K/F(T)$).

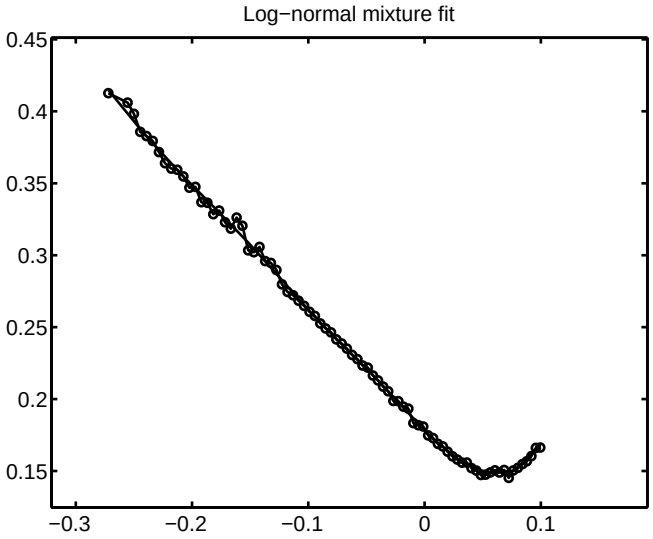


Figure 7: Filtered trajectories of the latent factor $\sqrt{v}$ when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over the S&P 500 log-returns, the VIX index values, VIX and S&P 500 option prices from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

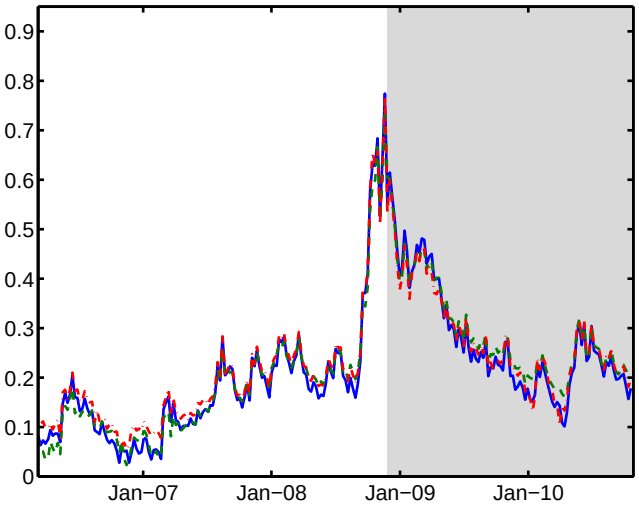


Figure 8: Difference between the filtered volatility when estimating the SVJ2 model to different datasets $j = 1, 2, 3$; *Dataset 4* being the benchmark. Panel A displays the difference with *Dataset 1*, Panel B with *Dataset 2* and Panel C with *Dataset 3*. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

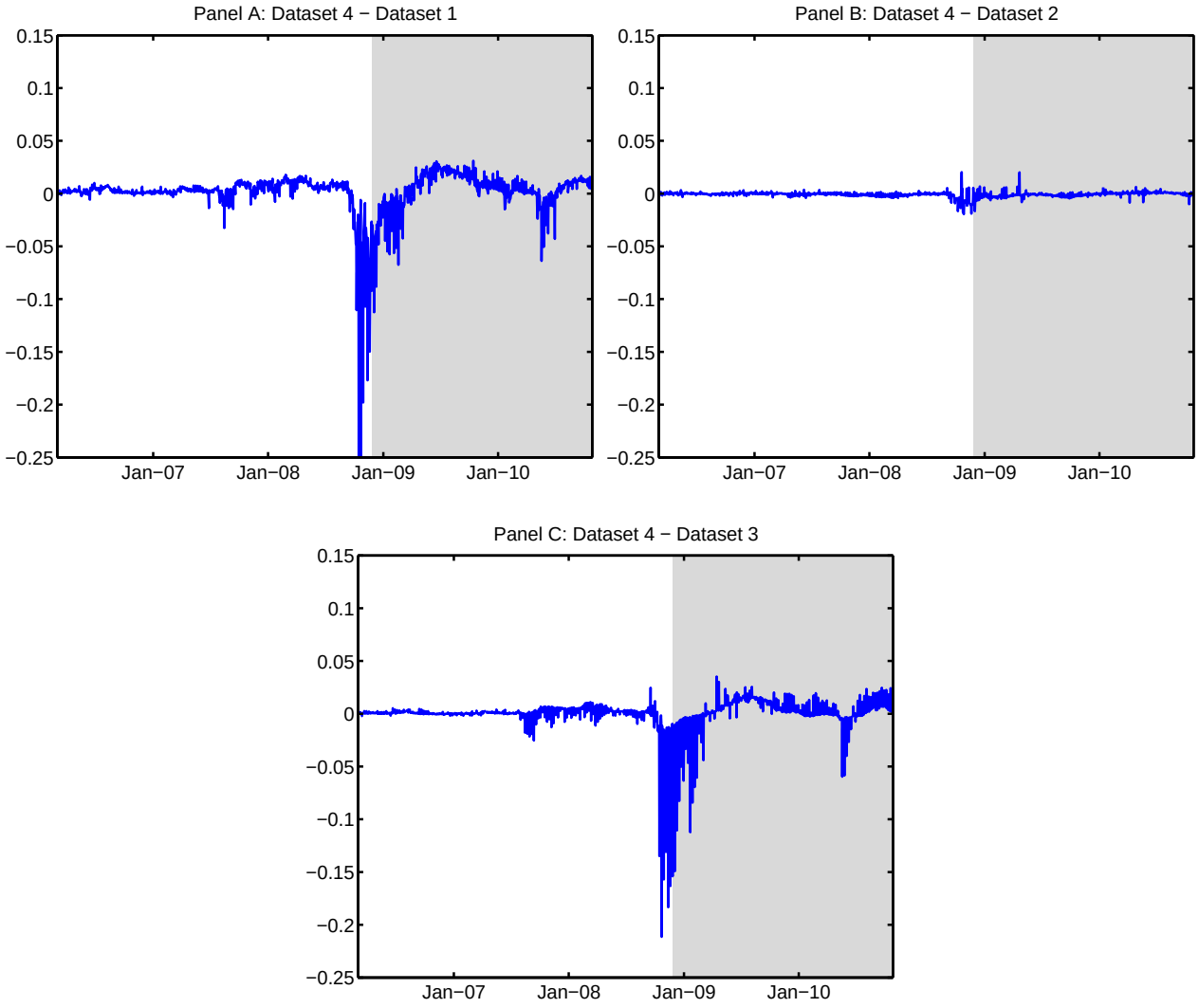


Figure 9: Filtered jump sizes in the variance process v when estimating the SVJ2 (solid line) and the SVJ (dashed line) models over the different datasets from March 2006 to November 2008 (685 days). We consider that there is a jump when the filtered probability of jump is larger than 50%. Panel A corresponds to *Dataset 1* which comprises the underlying forward returns on the S&P 500 and the VIX levels. Panel B corresponds to *Dataset 2* which consists of the underlying index levels plus S&P 500 options. Panel C corresponds to *Dataset 3* which comprises the underlying index levels plus VIX options. Finally Panel D corresponds to *Dataset 4* which gathers all data sources considered. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

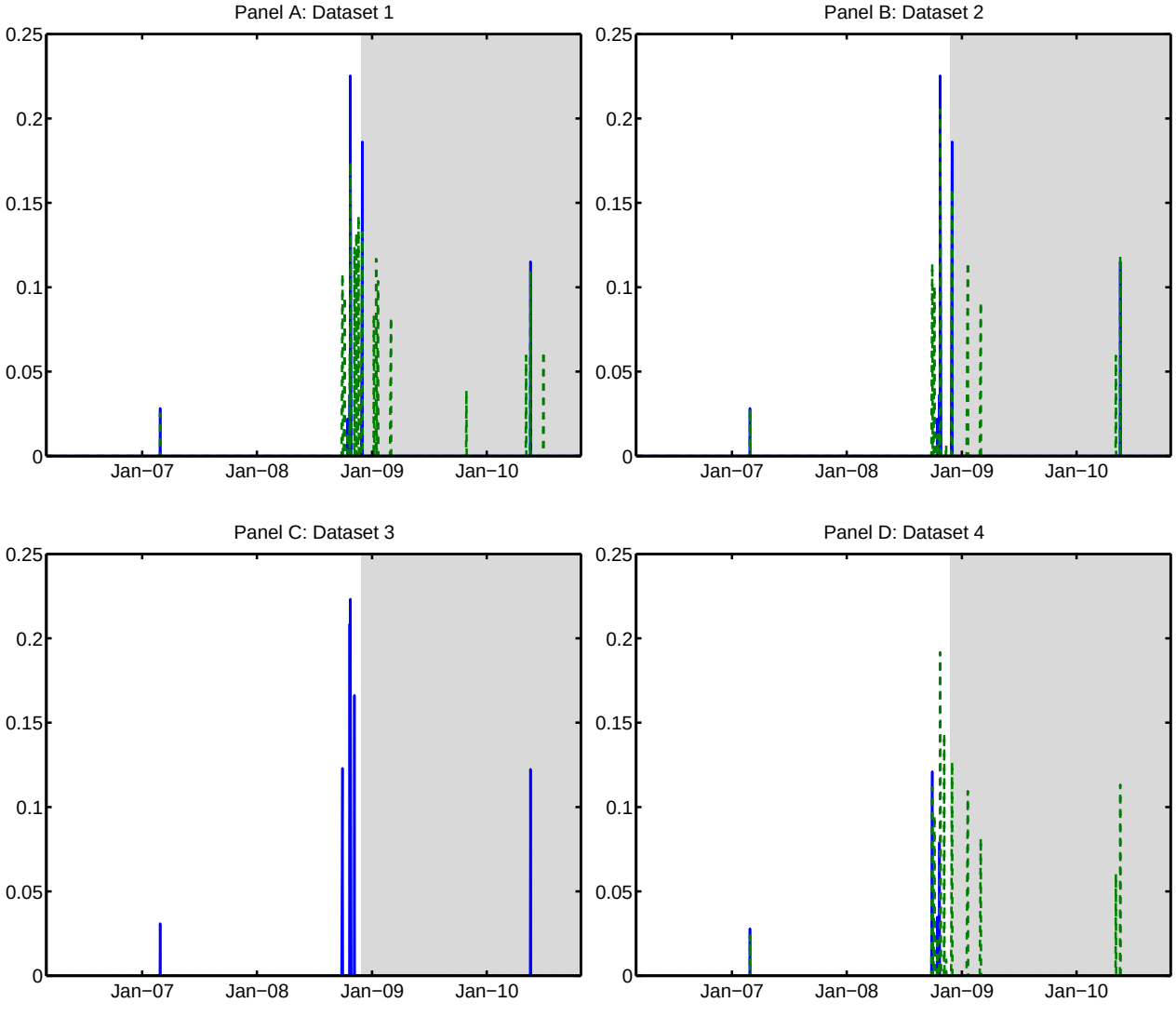


Figure 10: Filtered trajectories of the latent factor m when estimating the SVJ2 (solid line), SVJ (horizontal dashed line) and the SV2 (dashed dotted line) model over the different datasets from March 2006 to November 2008 (685 days). Each panel corresponds to a different dataset. *Dataset 1* is composed of SPX and VIX indices, *Dataset 2* of both indices and SPX options, *Dataset 3* of both indices and VIX options, *Dataset 4* of both indices and both SPX/VIX options. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

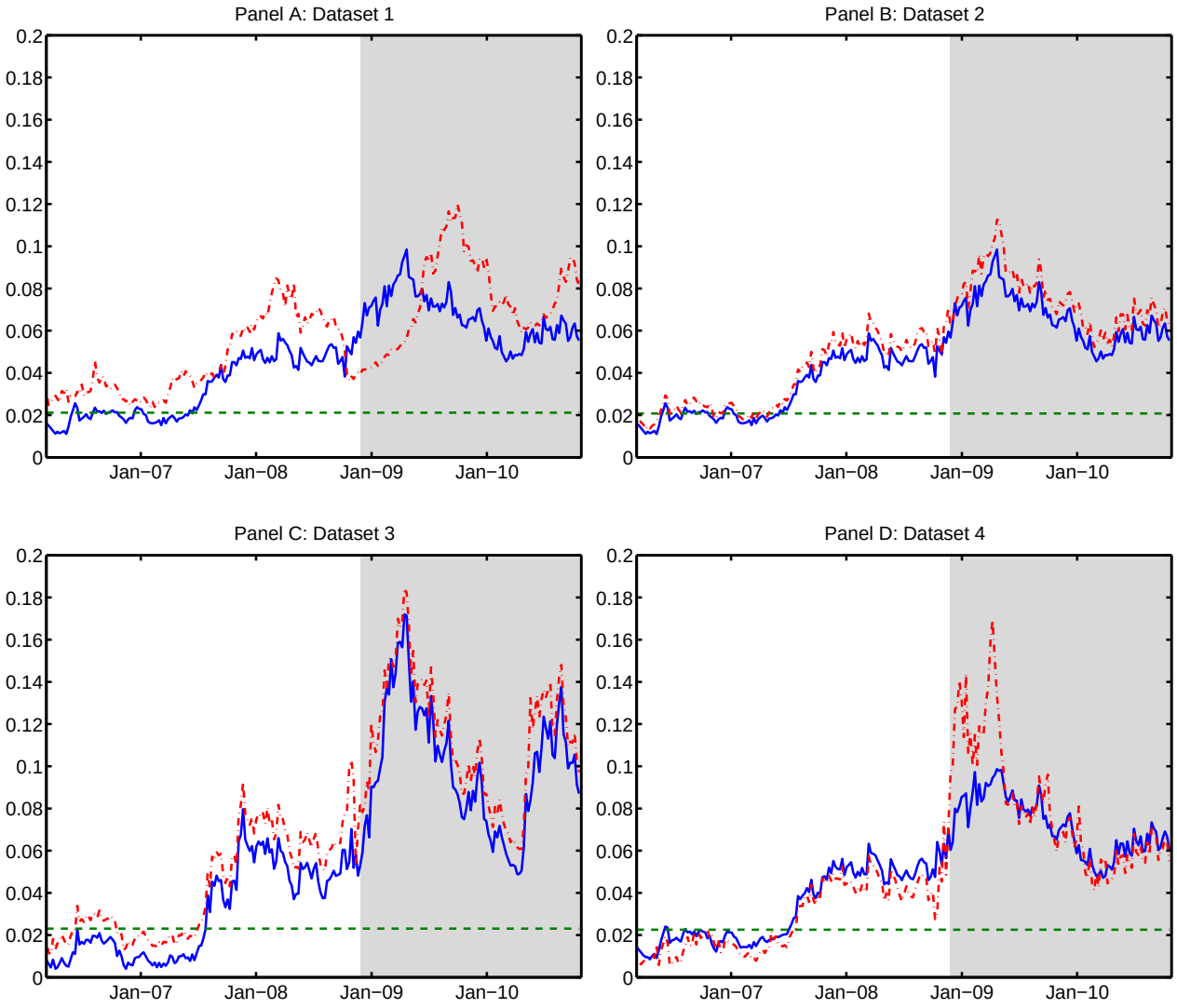


Figure 11: Fitting of the VIX index values for the SVJ2 model when the model is calibrated to log-returns and VIX levels (*Dataset 1*) from March 2006 to November 2008 (685 days). The crosses represent market data, the line the filtered values. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

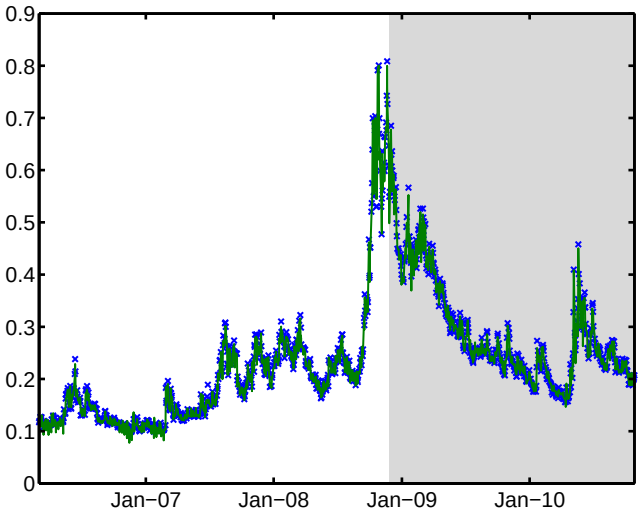


Figure 12: 1 month risk-neutral skewness and kurtosis of the distribution of returns implied by 1 month SPX options prices when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over *Dataset 4* (indices as well as SPX and VIX options) from March 2006 to November 2008 (685 days). We use the method described in Bakshi, Kapadia, and Madan (2003). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

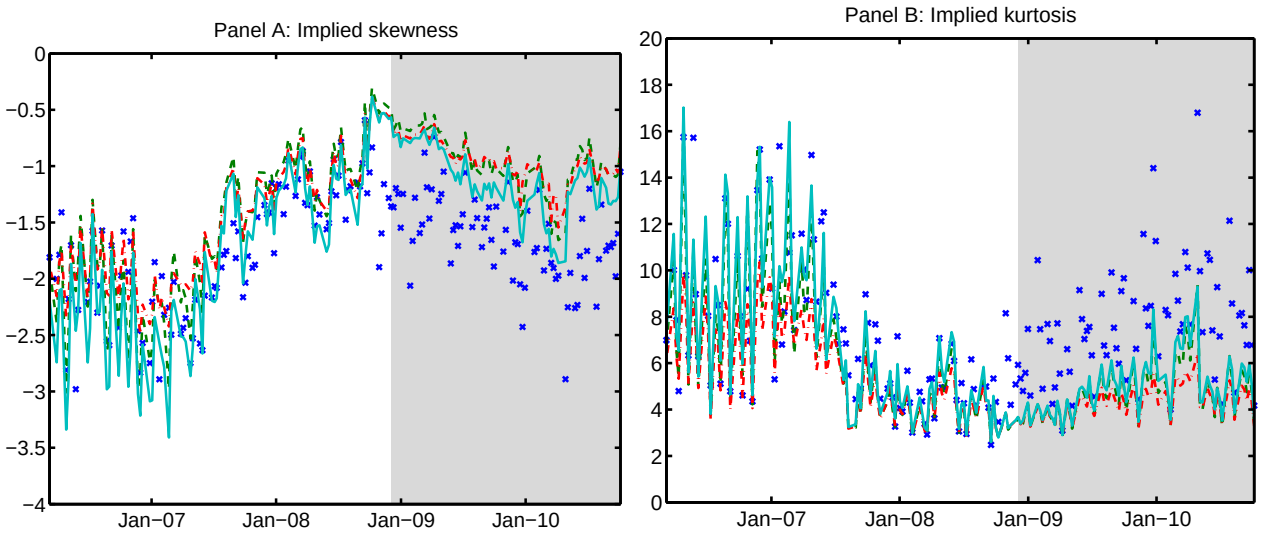


Figure 13: Filtered trajectories of the latent factor m when estimating the SVJ2 model over *Dataset 1* (solid line), *Dataset 2* (dashed line), *Dataset 3* (dashed-dotted line) and *Dataset 4* (dotted line) from March 2006 to November 2008 (685 days). *Dataset 1* is composed of SPX and VIX indices, *Dataset 2* of both indices and SPX options, *Dataset 3* of both indices and VIX options, *Dataset 4* of both indices and both SPX/VIX options. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

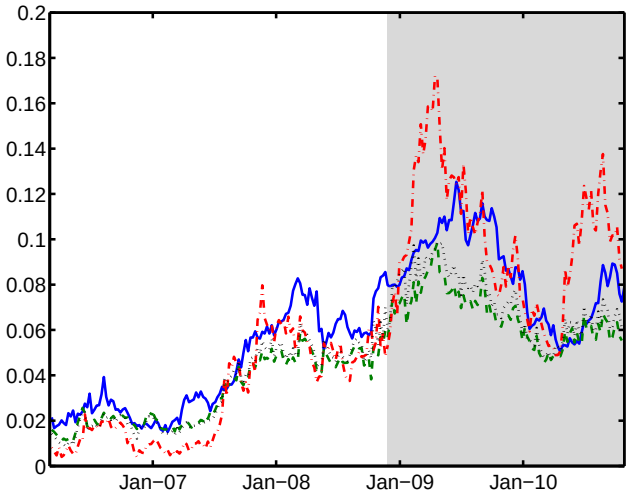


Figure 14: Integrated equity risk premium for different maturities when estimating the SVJ2 model using *Dataset 4* from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

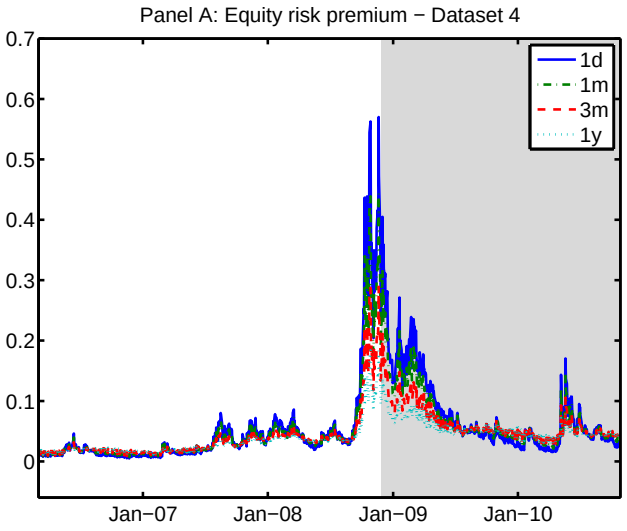
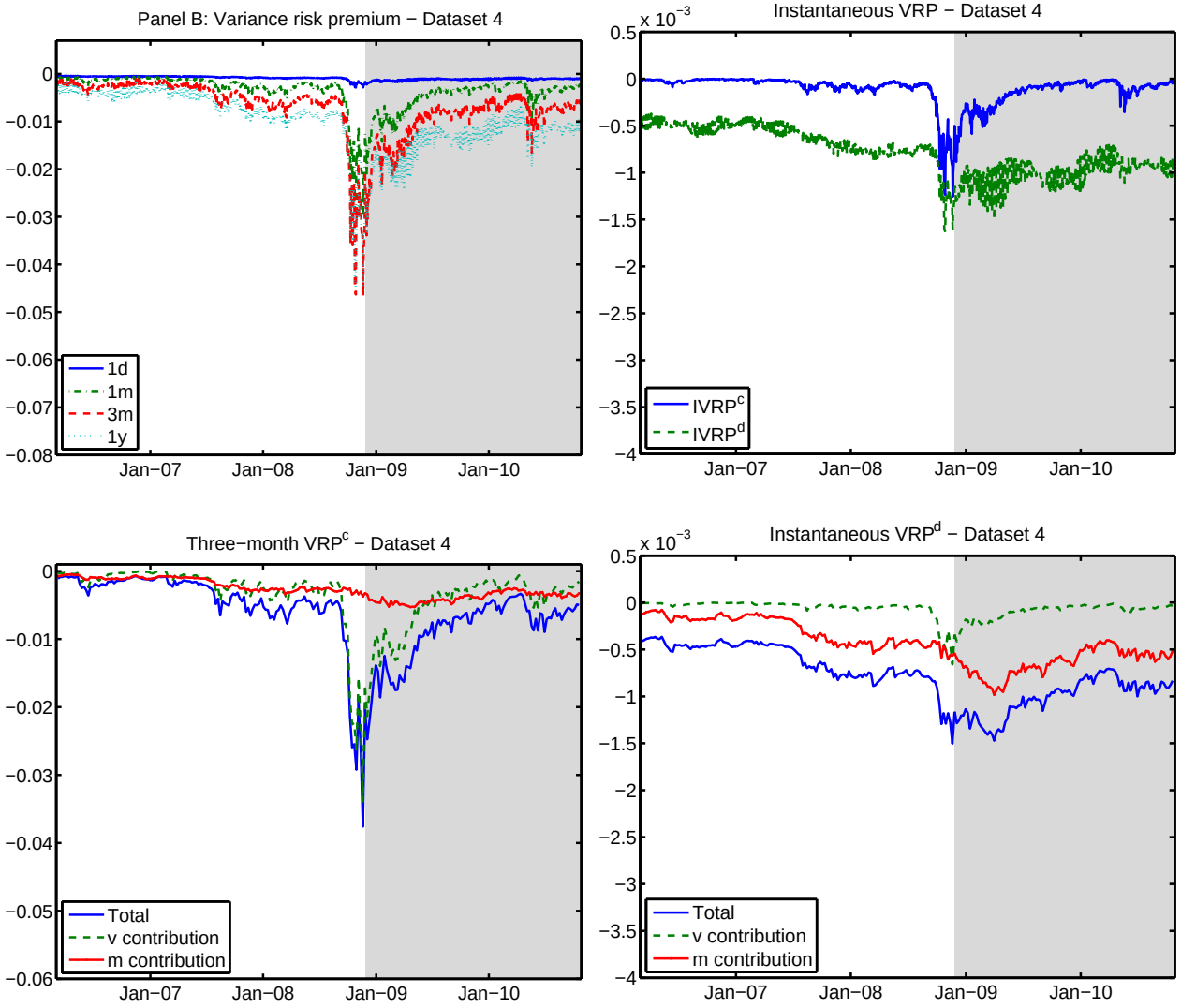


Figure 15: Panel A: Integrated variance risk premium for different maturities when estimating the SVJ2 model using *Dataset 4* from March 2006 to November 2008 (685 days). Panel B: Decomposition of the instantaneous IVRP into continuous and discontinuous components. Panel C: Contribution of the latent factors v and m in the continuous component of the 3-month IVRP. Panel D: Contribution of the latent factors v and m in the discontinuous component of the instantaneous IVRP. The shaded parts of the graphs represent the out-of-sample period, from 25 November 2008 until end of October 2010.



ONLINE APPENDIX TO

Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets

This appendix provides the results of technical derivations as well as the description of data of the data treatment before applying the particle filter, and the steps of the filter.

1 Model specification under $\mathbb{P}$

Under the historical measure $\mathbb{P}$, the model is specified as follows:

$$dY_t = [-\lambda^{Y(\mathbb{P})}(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{P})}(1,0,0) - 1) - \frac{1}{2}v_{t^-} + \gamma_t]dt + \sqrt{v_{t^-}}dW_t^{Y(\mathbb{P})} + dJ_t^{Y(\mathbb{P})} \tag{44}$$

$$dv_t = \kappa_v^{(\mathbb{P})}\left(\frac{\kappa_v^{(\mathbb{Q})}}{\kappa_v^{(\mathbb{P})}}m_{t^-} - v_{t^-}\right)dt + \sigma_v\sqrt{v_{t^-}}dW_t^{v(\mathbb{P})} + dJ_t^{v(\mathbb{P})} \tag{45}$$

$$dm_t = \kappa_m^{(\mathbb{P})}(\theta_m^{(\mathbb{P})} - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}dW_t^{m(\mathbb{P})} + dJ_t^{m(\mathbb{P})} \tag{46}$$

with:

$$\begin{aligned}
\gamma_t &= \eta_Y v_{t^-} + \lambda^Y(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{P})}(1,0,0) - \theta_Z^{(\mathbb{Q})}(1,0,0)) \\
dW_t^{Y(\mathbb{P})} &= dW_t^{Y(\mathbb{Q})} - \eta_Y\sqrt{v_{t^-}}dt \\
dW_t^{v(\mathbb{P})} &= dW_t^{v(\mathbb{Q})} + \sqrt{v_{t^-}}\frac{\kappa_v^{(\mathbb{P})} - \kappa_v^{(\mathbb{Q})}}{\sigma_v}dt \\
dW_t^{m(\mathbb{P})} &= dW_t^{m(\mathbb{Q})} + \sqrt{m_{t^-}}\frac{\kappa_m^{(\mathbb{P})} - \kappa_m^{(\mathbb{Q})}}{\sigma_m}dt \\
\theta_m^{(\mathbb{P})} &= \frac{\kappa_m^{(\mathbb{Q})}\theta_m^{(\mathbb{Q})}}{\kappa_m^{(\mathbb{P})}}.
\end{aligned}$$

2 Characteristic functions

The characteristic function of the processes Y, v and m defined in the model (1) - (3) are exponential affine as stated in Proposition 4.2.

$$\begin{aligned}
\Psi_{VIX_T^2}(t,v,m;\omega) &= \mathbb{E}_t\left[e^{\omega VIX_T^2}\right] = e^{\alpha(T-t)+\beta(T-t)v+\gamma(T-t)m}, \\
\Psi_{Y_T}(t,v,m;\omega) &= \mathbb{E}_t\left[e^{\omega Y_T}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)y+\gamma_Y(T-t)v+\delta_Y(T-t)m},
\end{aligned}$$

where $\omega \in \mathbb{C}$ is in each case chosen so that the integral converges.

The coefficients entering the definition of $\Psi_{VIX_T^2}$ satisfy the following ODEs:[32]

$$
\begin{aligned}
&-\alpha'(T-t)+\gamma(T-t)\kappa_m^{(\mathbb{Q})}\theta_m^{(\mathbb{Q})}+\lambda_0^{Yv}\left(\theta_Z^{(\mathbb{Q})}(0,\beta(T-t),0)-1\right)+\lambda_0^{m}\left(\theta_Z^{(\mathbb{Q})}(0,0,\gamma(T-t))-1\right)=0\\
&-\beta'(T-t)-\beta(T-t)\kappa_v^{(\mathbb{Q})}+\frac{1}{2}\sigma_v^2\beta^2(T-t)+\lambda_1^{Yv}\left(\theta_Z^{(\mathbb{Q})}(0,\beta(T-t),0)-1\right)=0\\
&-\gamma'(T-t)-\gamma(T-t)\kappa_m^{(\mathbb{Q})}+\frac{1}{2}\sigma_m^2\gamma^2(T-t)+\kappa_v^{(\mathbb{Q})}\beta(T-t)+\lambda_2^{Yv}\left(\theta_Z^{(\mathbb{Q})}(0,\beta(T-t),0)-1\right)+\\
&\qquad\lambda_1^{m}\left(\theta_Z^{(\mathbb{Q})}(0,0,\gamma(T-t))-1\right)=0
\end{aligned}
$$

$\forall t\in(0,T]$ with boundary conditions $\alpha(0)=0$, $\beta(0)=\omega_1$ and $\gamma(0)=\omega_2$, where $\omega_1:=\omega\alpha_{VIX^2}$ and $\omega_2:=\omega\beta_{VIX^2}$ (the coefficients α_{VIX^2} and β_{VIX^2} are defined in the Appendix 1).

The coefficients entering the definition of Ψ_{Y_T} satisfy the following ODEs are given by:

$$
\begin{aligned}
&-\alpha_Y'(T-t)+\beta_Y(T-t)(-\lambda_0^{Yv}(\theta_Z(\mathbb{Q})(1,0,0)-1))+\delta_Y(T-t)\kappa_m^{(\mathbb{Q})}\theta_m^{(\mathbb{Q})}\\
&\qquad+\lambda_0^{Yv}[\theta_Z^{(\mathbb{Q})}(\beta_Y(T-t),\gamma_Y(T-t),0)-1]+\lambda_0^{m}[\theta_Z^{(\mathbb{Q})}(0,0,\delta_Y(T-t))-1]=0\\
&-\beta_Y'(T-t)=0\\
&-\gamma_Y'(T-t)-\beta_Y(T-t)\lambda_1^{Yv}(\theta_Z^{(\mathbb{Q})}(1,0,0)-1)-\frac{1}{2}\beta_Y(T-t)-\gamma_Y(T-t)\kappa_v^{(\mathbb{Q})}+\frac{1}{2}\beta_Y(T-t)^2\\
&\qquad+\frac{1}{2}\gamma_Y(T-t)^2\sigma_v^2+\beta_Y(T-t)\gamma_Y(T-t)\sigma_v\rho_{Y,v}+\lambda_1^{Yv}[\theta_Z^{(\mathbb{Q})}(\beta_Y(T-t),\gamma_Y(T-t),0)-1]=0\\
&-\delta_Y'(T-t)-\beta_Y(T-t)\lambda_2^{Yv}(\theta_Z^{(\mathbb{Q})}(1,0,0)-1)+\gamma_Y(T-t)\kappa_v^{(\mathbb{Q})}-\delta_Y(T-t)\kappa_m^{(\mathbb{Q})}+\frac{1}{2}\delta_Y(T-t)^2\sigma_m^2\\
&\qquad+\lambda_2^{Yv}[\theta_Z^{(\mathbb{Q})}(\beta_Y(T-t),\gamma_Y(T-t),0)-1]+\lambda_1^{m}[\theta_Z^{(\mathbb{Q})}(0,0,\delta_Y(T-t))-1]=0
\end{aligned}
$$

$\forall t\in(0,T]$ with boundary conditions $\alpha_Y(0)=0$, $\beta_Y(0)=\omega$, $\gamma_Y(0)=0$ and $\delta_Y(0)=0$.

[32] This relies on the fact that the Poisson processes driving the jumps in v and in m are independent.

3 Coefficients of the VIX^2 formula

Table 11: Proposition 4.1 states that the VIX^2 depends on the instantaneous variance and level of mean reversion in an affine way. Here we give the values of coefficients playing a role in this proposition (see Appendix 1).

	A	B	G
$a_m \neq 0$ & $a_v \neq 0$ & $a_v \neq a_m$	$\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right)$	$\frac{1}{\tau_{\text{VIX}}} \frac{h_v}{(a_m - a_v)} \left[\left(\frac{e^{a_m \tau_{\text{VIX}}} - 1}{a_m}\right) - \left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v}\right)\right]$	$\frac{b_v}{a_v}\left[\left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v \tau_{\text{VIX}}}\right) - 1\right] - b_m B$
$a_m \neq 0$ & $a_v \neq 0$ & $a_v = a_m$	$\frac{1}{a \tau_{\text{VIX}}}\left(e^{a \tau_{\text{VIX}}} - 1\right)$	$\frac{h_v}{a}\left(e^{a \tau_{\text{VIX}}} - \frac{1}{a \tau_{\text{VIX}}}\left(e^{a \tau_{\text{VIX}}} - 1\right)\right)$	$\frac{b_v}{a}\left[\left(\frac{e^{a \tau_{\text{VIX}}} - 1}{a \tau_{\text{VIX}}}\right) - 1\right] - b_m B$
$a_m \neq 0$ & $a_v = 0$	1	$\frac{h_v}{a_m}\left(\frac{1}{a_m \tau_{\text{VIX}}}\left(e^{a_m \tau_{\text{VIX}}} - 1\right) - 1\right)$	$\frac{1}{2} b_v \tau_{\text{VIX}} - b_m B$
$a_m = 0$ & $a_v \neq 0$	$\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right)$	$\frac{h_v}{a_v}\left(\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right) - 1\right)$	$\frac{c_m}{a_v}\left(B - \frac{1}{2} h_v \tau_{\text{VIX}}\right)$ $+ \frac{1}{a_v} \frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0) \lambda_0^{Y^v}\left[\left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v \tau_{\text{VIX}}}\right) - 1\right]$
$a_m = 0$ & $a_v = 0$	1	$\frac{1}{2} \tau_{\text{VIX}} h_v$	$\frac{1}{2} \tau_{\text{VIX}}\left[\frac{\partial \theta_Z^{(\mathbb{Q})}}{\partial \phi_v}(0,0,0) \lambda_0^{Y^v} + c_m h_v \frac{\tau_{\text{VIX}}}{3}\right]$

	$\hat{A}$	$\hat{B}$
$a_m \neq 0$	$\frac{e^{a_m \tau_{\text{VIX}}} - 1}{a_m \tau_{\text{VIX}}}$	$b_m\left(1 - \hat{A}\right)$
$a_m = 0$	1	$\frac{c_m \tau_{\text{VIX}}}{2}$

4 Specific data treatment for the Particle filter

The dataset contains a large amount of ATM options compared to OTM and deep OTM options. This implies that if we use the filter (inside the maximum likelihood procedure) on this dataset and the model is not able to fit all options, the fitting of ATM options will be its priority rather than deep OTM options. Given the formula in Breeden and Litzenberger (1978), this results in fitting the body of the S&P 500 returns distribution rather than the tails which is not what we want: We need information about extreme events contained in the data to be incorporated into the models. For this reason, we interpolate the S&P 500 IV slices and re-sample option prices from the resulting parametric fit uniformly across moneynesses.[33] Other advantages of our interpolating method is that the resulting data is arbitrage free,[34] we have fewer points for each slice (but still representing accurately the information of each slice), thus reducing computational complexity.

We explain in detail in Appendix 5 how we have used the efficient mixture of log-normals approach

[33]It is common to interpolate data, see, e.g., Broadie, Chernov, and Johannes (2007). This eliminates arbitrage opportunities in the data and removes the accumulation of options around the ATM region.

[34]Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are a priori not guaranteed to be arbitrage free.

of Rebonato and Cardoso (2004) to have a parametric fit for each S&P 500 implied volatility slice. The root mean squared error of the S&P 500 implied volatilities parametric fits are on average around 0.25% and we therefore do not loose information especially given the market bid-ask spread. Finally, given the parametric fit for a given slice, we sample a fix number (we have chosen 15) of option prices uniformly distributed on the moneyness axis.

We do not perform any interpolation for the VIX options dataset as most VIX options are OTM and therefore contain information about the tails of the VIX distribution (i.e., variance and central tendency processes).

As the database comprises a large amount of options, it is unfeasible to calculate option prices every day for every particle, we follow Pan (2002) and Johannes, Polson, and Stroud (2009) among others and use weekly (Wednesday) option data. Furthermore, this eliminates beginning-of-week and end-of-week effects. Our filter uses daily time steps and incorporates information on the underlying indices on a daily basis.

The time-series of observations is decomposed into two periods, the first one from March 1st, 2006 to October 10, 2008 (before the peak of crisis and the VIX index increased to its highest point). This is a rather calm period,[35] that we will use as the in-sample estimation period. Our out-of-sample period starts on November 25, 2008 and ends on October 29, 2010. This period includes very high levels of volatility (i.e., implied volatilities from S&P 500 and VIX options as well as VIX index values). The last column of Table 4 reports the amount of options within each moneyness and maturity range in both periods.

[Insert Table 4 here]

In particular, our dataset contains 608 close-to-maturity OTM call options on the S&P 500 with moneyness larger than 1.05 and 2243 OTM put options with moneyness smaller than 0.95 in the in-sample period. These options have maturity shorter than two months. Analogously in the out-of-sample period the dataset comprises 737 close-to-maturity OTM call and 2032 OTM put options. As highlighted in Bollerslev and Todorov (2011), these options provide valuable information on jumps as they have small value unless a large movement in the S&P 500 occurs. Similarly, OTM calls on the VIX with short maturity contain information on the extreme upwards moves in the VIX index, and help identify the heavy-tailedness of the right tail of the VIX distribution. Our dataset comprises

[35] We have decided to include the beginning of October 2008 so that the in-sample period actually includes several dates with extreme events.

1006 such options on the VIX with moneyness larger than 1.1 in the in-sample period and 1269 in the out-of-sample period.

5 Interpolation procedure for the S&P 500 options dataset

The interpolation method we have used to find a parametric fit for each slice of S&P 500 implied volatilities is developed in detail in Rebonato and Cardoso (2004). Here we give the main idea and the particular choice of parameters we have made so that the slices are well fitted.

The idea of Rebonato and Cardoso (2004) is to use a mixture of log-normal densities for the Futures price.[36] Log-normal densities mixtures can fit various different shapes including multimodal densities arising from jumps in the Futures price process. The ability of the method to recover any type of density (or equivalently smile) is well documented in Rebonato and Cardoso (2004) and chapter 9 of Rebonato (2004).

The attractive feature of this parametric representation for the density of the Futures price is that the pricing of call/put options can be performed using a mixture of Black-Scholes price. Additionally, the no-arbitrage condition is simply a condition on the expectation for the mixture of the Futures price.

In practice, a mixture of 4 log-normal densities is enough to have a nearly perfect fit. We minimize the Euclidean distance between market and mixture option prices using the CMA-ES algorithm mentioned in section 5.1. We have checked that the resulting fits are satisfactory by computing different measures of the distance between the market and model implied volatility slices. For instance, the RMSE between implied volatilities of the parametric fit and the true implied volatilities is most of the time below 0.25%. Sometimes the RMSE is larger and goes up to 2%, however this is not due to the inability of the mixture of log-normals to fit an implied volatility slice but due to the shape of the data. This is best explained by looking at a typical fit as displayed on Figure 6. We can see that the fit is nearly perfect, however the RMSE is not so close to zero because the input data is too rough. This phenomenon is amplified if the data has a sawtooth pattern (potential arbitrage) although the fit is very good.

Finally, using the parametric fit, we can sample "market option prices" for the desired strikes and moneynesses. We have chosen to re-sample option prices from each parametric slice uniformly in

[36] They use a mixture for the stock price density but it is simple to adapt it to the Futures price when interest rates are constant.

strike. We also choose not to resample options for which the strike is smaller than 40% or larger than 140% of the current Futures price. The reason is that there are usually only one or two options outside this interval of moneyness and we do not wish to re-sample options where the interpolation results could be driven by an outlier.

6 Particle Filter

The filtering procedure consists in approximating the distribution $p(L_{t_k}|y^{t_k}, \Theta, \mathcal{M})$ of unobservable (latent) factors L_{t_k} at every point in time $t_k = k\Delta t$ given the available observations $y^{t_k} = \{y_{t_0}, ..., y_{t_k}\}$ and assuming that the model $\mathcal{M}$ is well specified and its parameters Θ are known. This approximation is denoted by $\hat{p}(L_{t_k}|y^{t_k}, \Theta, \mathcal{M})$ as in (39). In the remainder of this section, we drop the dependence on the parameter set Θ and the model specification $\mathcal{M}$. The available market measurements are the S&P 500 daily returns, the VIX levels and the option prices on both indices. We refer to section 5.2 for notations. The algorithm can be decomposed into the following steps.

(i) **Initialization**

We simulate n_p initial particles for the latent variables $\{v_{t_0}^{(i)}, m_{t_0}^{(i)}\}_{i=1,...,n_p}$ which are compatible with the initial value of the VIX squared, i.e., given the specification (28).

The steps (ii) to (v) described below are repeated for each time step t_k in the grid from $k = 0$ to $k = M - 1$.

(ii) **First-stage resampling**

At this point, we assume that we have n_p particles (i.e., possible values of m and v) at time t_k given all observations y^{t_k} up to t_k. At time t_{k+1}, there are new observations y_{k+1}. The goal of this step is to keep from the previous sample of particles $\{v_{t_k}^{(i)}, m_{t_k}^{(i)}\}_{1\leq i\leq n_p}$ only those which are likely to generate the new observation y_{k+1}. For this purpose, we assign a weight (namely *first-stage weights*) to each particle which is proportional to the likelihood of new market observations $y_{t_{k+1}}$ given the value of the particle L_{t_k} at time t_k. Intuitively, particles that are compatible with the new observations will be assigned larger weights than other particles. To increase the speed of the first-stage resampling, we do not consider options as part of the observations $y_{t_{k+1}}$ (only in this step) and limit $y_{t_{k+1}}$ to the values of the indices.

The first-stage weight $\omega_{t_{k+1}}^{(i)}$ assigned to the i^{th} particle $L_{t_k}^{(i)}$ at time t_{k+1} is given by:

$$\omega_{t_{k+1}}^{(i)} = p(L_{t_k}^{(i)}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|L_{t_k}^{(i)})$$

where $p(y_{t_{k+1}}|L_{t_k}^{(i)})$ is the density of the observation vector $y_{t_{k+1}}$ given the values of the particle vector $L_{t_k}^{(i)}$. The importance weights $\{\omega_{t_{k+1}}^{(i)}\}_{1\leq i\leq n_p}$ add up to 1 so that they define a proper probability distribution. Conditioning on the number of jumps in Y (or v) and m gives:

$$\omega_{t_{k+1}}^{(i)} \propto \sum_{j,l\in\mathbb{N}} p(y_{t_{k+1}}|L_{t_k}^{(i)}, \Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m})\mathbb{P}(\Delta N_{t_k}^{Yv} = j, \Delta N_{t_k}^{m} = l).$$

Given that we use daily observations, we limit the possible number of jumps of the Poisson random variables $\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}$ to zero or one (Bernoulli approximation). We recall that the new observation is composed of the SPX returns and the VIX level $y_{t_{k+1}} = (\Delta Y_{t_{k+1}}, VIX_{t_{k+1}}^2)$. Since the $VIX_{t_{k+1}}^2$ is the sum of normal distributions and *up to two* exponential distributions, there is no closed form for this bivariate density in the general case. To preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (generalization of a Bernoulli distribution) which has support a certain number of (the corresponding exponential distribution) quantiles. As a consequence, the weight $\omega_{t_{k+1}}^{(i)}$ is a sum of weighted bivariate normal densities.

To eliminate the particles $\{L_{t_k}^{(i)}\}_{1\leq i\leq n_p}$ that are not likely to generate the new observations $y_{t_{k+1}}$, we resample (with replacement) particles according to a stratified resampling scheme:[37]

$$z(i) \sim StratRes(n_p, \omega_{t_{k+1}}^{(1)}, ..., \omega_{t_{k+1}}^{(n_p)}).$$

This makes it possible to create a new sample of n_p latent factors $\{L_{t_k}^{(j)}\}_{1\leq j\leq n_p}$ which are now equally likely. Indeed, particles representing m_{t_k} and v_{t_k} are shuffled into a new set of particles: $\{m_{t_k}^{(j)}, v_{t_k}^{(j)}\}_{j=1..n_p} = \{m_{t_k}^{z(i)}, v_{t_k}^{z(i)}\}_{i=1..n_p}$. We resample the same number of particles although this is in principle not necessary.

The next step of the particle filter consists in propagating the latent factors according to their

[37] We checked that using a multinomial or stratified resampling scheme gives the same results.

conditional density given the previous values $L_{t_k}^{(i)}$ and the new observations $y_{t_{k+1}}$:

$$L_{t_{k+1}}^{(i)} \sim p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}}).$$

Because the distribution $p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$ is not known in closed-form, we use a proposal density $q(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$. Propagating v and m requires preliminary knowledge on the jump components so we first focus on the jumps.

(iii) **Generating the jumps**

We calculate the joint probability of jumps in Y (or v) and m between t_k and t_{k+1} using:

$$\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m})\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}). \tag{47}$$

Conditionally on the jump sizes in v and m, the first part of the right hand-side has already been calculated in the first-stage weights. Using Bayes' rule, we get an approximation for $\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}|y_{t_{k+1}})$. However, to have a better chance at detecting extreme events, we force the probability of jumps to be at least 10% for all processes and simulate from the resulting distribution function the jumps.

We infer the jump size in the returns following Johannes, Polson, and Stroud (2009):

$Z_{t_k}^{Y(i)}|\Delta N_{t_k}^{Yv}, y_{t_{k+1}}$ is normally distributed $\mathcal{N}(\mu_J^{Y(i)}, \sigma_J^{Y(i)})$ where $\mu_J^{Y(i)}$ and $\sigma_J^{Y(i)}$ are given by

$$\begin{aligned}(\sigma_J^{Y(i)})^2 &= \left(\frac{1}{\Delta N_{t_k}^{Yv(i)}\sigma_Y^2} + \frac{1}{\hat{v}_{t_{k+1}}^{(i)}}\right)^{-1} \\ \mu_J^{Y(i)} &= (\sigma_J^{Y(i)})^2 \frac{Y_{t_{k+1}} - \tilde{\mu}_Y^{(i)}}{\hat{v}_{t_{k+1}}^{(i)}} + \frac{(\sigma_J^{Y(i)})^2}{\sigma_Y^2}\mu_Y\end{aligned}$$

where $\hat{v}_{t_{k+1}}^{(i)}$ is an estimate of $v_{t_{k+1}}$ given the information we have up to time t_k and particle i; we use $\hat{v}_{t_{k+1}}^{(i)} = \mathbb{E}[v_{t_{k+1}}|v_{t_k}^{(i)}]$ and

$$\tilde{\mu}_Y^{(i)} = -\left(\lambda^{Yv}(\theta_Z^{\mathbb{Q}}(1,0,0) - 1) + \frac{1}{2}v_{t_k}^{(i)} - \Delta N_{t_k}^{Yv(i)}\right)\Delta t.$$

Finally, we simulate jump sizes for v and m according to their exponential law.

(iv) **Propagating the volatility and central tendency**

The latent factors v and m are propagated following a Milstein discretization scheme of the SDE. We use the full truncation method to prevent them from taking negative values.

(v) **Computing the filtering density**

At this point, the newly generated particles $\{L^{(i)}_{t_{k+1}}\}_{1\leq i\leq n_p}$ are a sample of $p(L_{t_{k+1}}|y^{t_{k+1}})$. We now calculate the *second-stage weights* $\{\pi^{(i)}_{t_{k+1}}\}_{1\leq i\leq n_p}$ which approximate the probabilities $p(L^{(i)}_{t_{k+1}}|y^{t_{k+1}})$ and give an approximation for the filtering density at time t_{k+1}. These weights are proportional to the likelihood of observations at time t_{k+1} given the propagated particles $L^{(i)}_{t_{k+1}}$, with a correction related to the proposal density:

$$\pi^{(i)}_{t_{k+1}} \propto \frac{p(L^{(i)}_{t_{k+1}}|L^{(i)}_{t_k})p(y_{t_{k+1}}|L^{(i)}_{t_{k+1}})}{\omega^{z(i)}_{t_{k+1}} q(L^{(i)}_{t_{k+1}}|L^{(i)}_{t_k}, y_{t_{k+1}})}.$$

The posterior distribution of state variables is approximated by:

$$\hat{p}(L_{t_{k+1}}|y^{t_{k+1}}) = \sum_{i=1}^{n_p} \pi^{(i)}_{t_{k+1}} \delta(L_{t_{k+1}} - L^{(i)}_{t_{k+1}})$$

.

We choose the most-likely value of a given factor by taking the expectation of the estimated filtering density, e.g., $\hat{v}_{t_{k+1}} = \mathbb{E}_{\hat{p}}[v^{(i)}_{t_{k+1}}]$.

The algorithm described above extracts latent factors if one assumes that the model parameters are known. Pitt (2002) builds on Gordon, Salmond, and Smith (1993) to show that the parameters can be estimated using the Maximum Likelihood Importance Sampling Criterion, defined as the product over time of the averages of the second-stage weights. The likelihood of observations given the values of particles is then estimated by the average of the second-stage weights over particles:

$$\hat{p}(y^{t_n}|\Theta, \mathcal{M}) = \prod_{k=1}^{M} \hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M})\hat{p}(y_{t_0}|\Theta, \mathcal{M})$$

where

$$\hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) = \frac{1}{n_p}\sum_{i=1}^{n_p} \pi^{(i)}_{t_k}.$$

7 Decomposition of risk premia

We respectively denote by IERP^c and IERP^d the continuous and discontinuous components of the IERP:

$$IERP(t,T) = IERP^c(t,T) + IERP^d(t,T)$$

with

$$IERP^c(t,T) = \frac{1}{T-t}\eta_Y \mathbb{E}_t^{\mathbb{P}}\left[\int_t^T v_{s^-} ds\right]$$

and

$$\begin{aligned} IERP^d(t,T) =& \frac{1}{T-t}\Big(\int_t^T \mathbb{E}_t^{\mathbb{P}}[\exp(Z_s^{Y(\mathbb{P})}-1)dN_s^{Yv} - \lambda^{Yv}(v_{s^-},m_{s^-})(\theta_Z^{(\mathbb{Q})}(1,0,0)-1)ds] \\ & - \mathbb{E}_t^{\mathbb{Q}}[\exp(Z_s^{Y(\mathbb{Q})}-1)dN_s^{Yv} - \lambda^{Yv}(v_{s^-},m_{s^-})(\theta_Z^{(\mathbb{Q})}(1,0,0)-1)ds]\Big) \end{aligned}$$

As the jump intensity is an affine function of the latent factors v and m, all calculations boil down to a linear combination of $\int_t^T \mathbb{E}_t^{\mathbb{P}}[v_{s^-}]ds$ and $\int_t^T \mathbb{E}_t^{\mathbb{P}}[m_{s^-}]ds$, which are known in closed form. Indeed, extending the coefficients in Table 11 to functions $A(\tau)$, $B(\tau)$, $G(\tau)$ of a variable time-to-maturity τ, we have:

$$\mathbb{E}_t^{\mathbb{P}}[m_{u^-}] = \alpha_m^{\mathbb{P}}(t,u)m_t + \beta_m^{\mathbb{P}}(t,u)$$

for some functions $\alpha_m^{\mathbb{P}}$ and $\beta_m^{\mathbb{P}}$ which are solutions of ODEs and have already been calculated as part of the computation of the VIX squared coefficients. Similarly for the expectation of the integrated variance:

$$\mathbb{E}_t^{\mathbb{P}}[v_{u^-}] = \alpha_v^{\mathbb{P}}(t,u)v_t + \beta_v^{\mathbb{P}}(t,u)\frac{\kappa_v^{\mathbb{Q}}}{\kappa_v^{\mathbb{P}}}m_t + \gamma_v(t,u).$$

for some functions α_v, β_v and γ_v.

Hence:

$$\int_t^T \mathbb{E}_t^{\mathbb{P}}[m_{s^-}]ds = (T-t)(\hat{A}^{\mathbb{P}}(T-t)m_t + \hat{B}^{\mathbb{P}}(T-t))$$

and

$$\int_t^T \mathbb{E}_t^{\mathbb{P}}[v_{s^-}]ds = (T-t)(A^{\mathbb{P}}(T-t)v_t + B^{\mathbb{P}}(T-t)\frac{\kappa_v^{\mathbb{Q}}}{\kappa_v^{\mathbb{P}}}m_t + G^{\mathbb{P}}(T-t)).$$

Therefore the continuous part of the IERP at every time $t \geq 0$ can expressed as a linear combination of v_t and m_t:

$$IERP^c(t,T) = \alpha_{IERP^c}(t,T)v_t + \beta_{IERP^c}(t,T)m_t + \gamma_{IERP^c}(t,T).$$

The functions α_{IERP^c}, β_{IERP^c} and γ_{IERP^c} are given by:

$$\alpha_{IERP^c}(t,T) = \eta_Y A^{\mathbb{P}}(T-t),$$

$$\beta_{IERP^c}(t,T) = \eta_Y B^{\mathbb{P}}(T-t)\frac{\kappa_v^{\mathbb{Q}}}{\kappa_v^{\mathbb{P}}},$$

$$\gamma_{IERP^c}(t,T) = \eta_Y G^{\mathbb{P}}(T-t).$$

Similarly, the discontinuous IERP can be decomposed into:

$$IERP^d(t,T) = \alpha_{IERP^d}(t,T)v_t + \beta_{IERP^d}(t,T)m_t + \gamma_{IERP^d}(t,T)$$

where the coefficients α_{IERP^d}, β_{IERP^d} and γ_{IERP^d} are given by:

$$\alpha_{IERP^d} = \lambda_1^{Yv} A^{\mathbb{P}}(T-t)(\theta_Z^{\mathbb{P}}(1,0,0) - \theta_Z^{\mathbb{Q}}(1,0,0)),$$

$$\beta_{IERP^d} = \left(\lambda_1^{Yv} B^{\mathbb{P}}(T-t)\frac{\kappa_v^{\mathbb{Q}}}{\kappa_v^{\mathbb{P}}} + \lambda_2^{Yv}\hat{A}^{\mathbb{P}}(T-t)\right)(\theta_Z^{\mathbb{P}}(1,0,0) - \theta_Z^{\mathbb{Q}}(1,0,0)),$$

$$\gamma_{IERP^d} = (\lambda_0^{Yv} + \lambda_1^{Yv} G^{\mathbb{P}}(T-t) + \lambda_2^{Yv}\hat{B}^{\mathbb{P}}(T-t))(\theta_Z^{\mathbb{P}}(1,0,0) - \theta_Z^{\mathbb{Q}}(1,0,0)).$$

The same procedure yields to a similar decomposition of the IVRP:

$$IVRP(t,T) = IVRP^c(t,T) + IVRP^d(t,T)$$

with

$$IVRP^c(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T v_{s^-}ds\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T v_{s^-}ds\right)\right],$$

$$IVRP^d(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\sum_{t\le s\le T}(J_s^Y)^2\Delta N_s\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\sum_{t\le s\le T}(J_s^Y)^2\Delta N_s\right)\right].$$

Each part can also be decomposed into a contribution of m and another one of v as follows:

$$IVRP^c(t,T) = \alpha_{IVRP^c}(t,T)v_t + \beta_{IVRP^c}(t,T)m_t + \gamma_{IVRP^c}(t,T)$$

with

$$\alpha_{IVRP^c}(t,T) = A^{\mathbb{P}}(T-t) - A^{\mathbb{Q}}(T-t),$$

$$\beta_{IVRP^c}(t,T) = B^{\mathbb{P}}(T-t)\frac{\kappa_v^{\mathbb{Q}}}{\kappa_v^{\mathbb{P}}} - B^{\mathbb{Q}}(T-t),$$

$$\gamma_{IVRP^c}(t,T) = G^{\mathbb{P}}(T-t) - G^{\mathbb{Q}}(T-t).$$

$$IVRP^d(t,T) = \alpha_{IVRP^d}(t,T)v_t + \beta_{IVRP^d}(t,T)m_t + \gamma_{IVRP^d}(t,T)$$

with

$$\alpha_{IVRP^d}(t,T) = \lambda_1^{Yv}\left[(\sigma_Y^{\mathbb{P}})^2 + (\mu_Y^{(\mathbb{P})})^2 - (\sigma_Y^{\mathbb{Q}})^2 - (\mu_Y^{(\mathbb{Q})})^2\right],$$

$$\beta_{IVRP^d}(t,T) = \lambda_2^{Yv}\left[(\sigma_Y^{\mathbb{P}})^2 + (\mu_Y^{(\mathbb{P})})^2 - (\sigma_Y^{\mathbb{Q}})^2 - (\mu_Y^{(\mathbb{Q})})^2\right],$$

$$\gamma_{IVRP^d}(t,T) = \lambda_0^{Yv}\left[(\sigma_Y^{\mathbb{P}})^2 + (\mu_Y^{(\mathbb{P})})^2 - (\sigma_Y^{\mathbb{Q}})^2 - (\mu_Y^{(\mathbb{Q})})^2\right].$$

REFERENCES

Aït-Sahalia, Y., M. Karaman, and L. Mancini, 2012, "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis," working paper, Swiss Finance Institute Working Paper.

Aït-Sahalia, Y., and R. Kimmel, 2007, "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

Aït-Sahalia, Y., and A. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53, 500ï¿½–547.

Alizadeh, S., M. Brandt, and M. Diebold, 2002, "Range-based estimation of stochastic volatility models," *The Journal of Finance*, 57, 1047–1091.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.

Andrews, D., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

Bakshi, G., N. Kapadia, and D. Madan, 2003, "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *The Review of Financial Studies*, 16, 101–143.

Bates, D. S., 1996, "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

——— , 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

——— , 2003, "Empirical option pricing: a Retrospection," *Journal of Econometrics*, 116, 387 – 404.

——— , 2012, "U.S. stock market crash risk, 1926-2010," *Journal of Financial Economics*, 105(2), 229 – 259.

Bayer, C., J. Gatheral, and M. Karlsmark, 2013, "Fast Ninomiya-Victoir calibration of the Double-Mean-Reverting Model," working paper.

Bergomi, L., 2009, "Smile Dynamics IV," *Risk*, Dec., 90 – 96.

Bollerslev, T., and V. Todorov, 2011, "Tails, Fears and Risk Premia," *Journal of Finance*, 66, 2165–2211.

Breeden, D., and R. H. Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

Broadie, M., M. Chernov, and M. Johannes, 2007, "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 62(3), 1453–1490.

Carr, P., and D. B. Madan, 1999, "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, 2(4), 1–18.

Carr, P., and L. Wu, 2009, "Variance risk premiums," *Review of Financial Studies*, 22, 1311–1341.

CBOE, C. B. O. E., 2009, "THE CBOE VOLATILITY INDEX - VIX," working paper, Chicago Board Options Exchange, White paper.

Chernov, M., A. R. Gallant, E. Ghysels, and G. T. Tauchen, 2003, "Alternative Models of Stock Price Dynamics," *Journal of Econometrics*, 116, 225 – 257.

Chernov, M., and E. Ghysels, 2000, "A Study towards a Unified Approach to the Joint Estimation of Objective ad Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407–458.

Christoffersen, P., and K. Jacobs, 2004, "The importance of the loss function in option valuation," *Journal of Financial Economics*, 72(2), 291 – 318.

Christoffersen, P., K. Jacobs, and K. Mimouni, 2010, "Models for S&P 500 Dynamics: Evidence from Realized Volatility, Daily Returns and Options Prices," *Review of Financial Studies*, 23(8), 3141–3189.

Chung, S., W. Tsai, Y. Wang, and P. Wenig, 2011, "The Information Content of the S&P 500 Index and VIX Options on the Dynamics of the S&P 500 Index," *The Journal of Futures Markets*, 31(12), 1170 – 1201.

Cont, R., and T. Kokholm, 2011, "A Consistent Pricing Model for Index Options and Volatility Derivatives," *Mathematical Finance*.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou, 1999, "More than you ever wanted to know about volatility swaps," *Goldman Sachs Quantitative Strategies Research Notes*.

Detemple, J., and C. Osakwe, 2000, "The valuation of volatility options," *European Finance Review*, 4, 21–50.

Drimus, G. G., and E. W. Farkas, 2013, "Local Volatility of Volatility for the VIX Market," *Review of Derivatives Research*, (735), To appear.

Duan, J.-C., and C.-Y. Yeh, 2011, "Price and Volatility Dynamics Implied by the VIX Term Structure ," working paper, NUS RMI Working Paper No. 11/05.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Durham, G. B., 2012, "Risk-neutral modelling with affine and non-affine models," .

Egloff, D., M. Leippold, and L. Wu, 2010, "Valuation and Optimal Investing in Variance Swaps," *Journal of Financial and Quantitative Analysis*, 45(5), 1279–1310.

Eraker, B., 2004, "Do stock prices and volatility jump? Reconciling evidence from spot and option prices," *Journal of Finance*, 59, 1367–1403.

Eraker, B., M. S. Johannes, and N. Polson, 2003, "The Impact of Jumps in Equity Index Volatility

and Returns," *Journal of Finance*, 58, 1269–1300.

Fang, F., and C. Oosterlee, 2008, "A novel pricing method for European options based on Fourier-cosine series expansions," working paper, MPRA Paper 9319, University Library of Munich, Germany.

Ferriani, F., and S. Pastorello, 2012, "Estimating and Testing Non-Affine Option Pricing Models With a Large Unbalanced Panel of Options," *The Econometrics Journal*, 15(2), 171–203.

Gatheral, J., 2008, "Consistent Modeling of SPX and VIX options," The Fifth World Congress of the Bachelier Finance Society, London, July 2008.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, 140(2), 107–113.

Grünbichler, A., and F. Longstaff, 1996, "Valuing futures and options on volatility," *Journal of Banking and Finance*, 20, 985–1001.

Hansen, N., and A. Ostermeier, 1996, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC '96)*, pp. 312–317.

Jackwerth, J., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

Jacod, J., and V. Todorov, 2010, "Do Price and Volatility Jump Together?," *Annals of Applied Probability*, 20, 1425–1469.

Jiang, G. J., and Y. S. Tian, 2007, "Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index," *The Journal of Derivatives*, 14, 35–60.

Johannes, M. S., N. G. Polson, and J. R. Stroud, 2009, "Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices," *The Review of Financial Studies*, 22(7), 2759–2799.

Jones, C. S., 2003, "The Dynamics of Stochastic Volatility: Evidence From Underlying and Options Markets," *Journal of Econometrics*, 116, 181–224.

Kaeck, A., and C. Alexander, 2012, "Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions," *Journal of Banking & Finance*, 36(11), 3110 – 3121.

Lian, G.-H., and S.-P. Zhu, 2011, "Pricing VIX options with stochastic volatility and random jumps," *Decisions in Economics and Finance*, pp. 1–18.

Lindström, E., J. Ströjby, M. Brodén, M. Wiktorsson, and J. Holst, 2008, "Sequential calibration of options," *Computational Statistics & Data Analysis*, 52(6), 2877–2891.

Mencía, J., and E. Sentana, 2013, "Valuation of VIX derivatives," *Journal of Financial Economics*, 108(2), 367 – 391.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and

Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pan, J., 2002, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.

Papanicolaou, A., and R. Sircar, 2012, "A Regime-Switching Heston Model for VIX and S&P 500 Implied Volatilities," *Quantitative Finance*, Forthcoming.

Pitt, M. K., 2002, "Smooth particle filters for likelihood evaluation and maximisation," working paper.

Pitt, M. K., and N. Shephard, 1999, "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446).

Rebonato, R., 2004, *Volatility and correlation*. Wiley, 2. ed. edn.

Rebonato, R., and T. Cardoso, 2004, "Unconstrained Fitting of Implied Volatility Surfaces Using a Mixture of Normals," *Journal of Risk*, 7(1), 55–74.

Ruijter, M., M. Versteegh, and C. W. Oosterlee, 2013, "On the Application of Spectral Filters in a Fourier Option Pricing Technique," working paper, SSRN, Working Paper.

Sepp, A., 2008a, "Pricing Options on Realized Variance in the Heston Model with Jumps in Returns and Volatility," *Journal of Computational Finance*, 11(4), 33–70.

——— , 2008b, "VIX Option Pricing in a Jump-Diffusion Model," *Risk Magazine*, pp. 84–89.

Song, Z., and D. Xiu, 2012, "A Tale of Two Option Markets: State-Price Densities Implied from S&P 500 and VIX Option Prices," working paper, Chicago Booth Research Paper No 12-10 - Fama-Miller Working Paper.

Storn, R., 1996, "On the usage of differential evolution for function optimization," in *Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, pp. 519–523.

Todorov, V., 2010, "Variance Risk Premium Dynamics: The Role of Jumps," *The Review of Financial Studies*, 23(1), 345–383.

Todorov, V., and G. Tauchen, 2011, "Volatility Jumps," *Journal of Business and Economic Statistics*, 29, 356–371.

Whaley, R. E., 1993, "Derivatives on Market Volatility: Hedging Tools Long Overdue," *The Journal of Derivatives*, 1, 71–84.

Wu, L., 2011, "Variance Dynamics: Joint Evidence from Options and High-Frequency Returns," *Journal of Econometrics*, 160, 280–287.

A parsimonious stochastic correlation framework to model the joint dynamics of assets

Chris Bardgett

Abstract

We introduce a novel and flexible stochastic correlation framework for asset returns. The system of stochastic differential equations satisfied by the correlation processes is proven to admit a unique strong solution and the correlation matrix is shown to be positive semidefinite over time. We investigate the case of correlated returns each of them following a stochastic volatility model and argue that our setup presents two main advantages compared to existing ones. First, the stochastic dependence structure is specified independently of the asset's individual dynamics, which makes it possible to estimate separately each asset's dynamics and their dependence structure. Second, our framework is parsimonious in the number of stochastic factors which is proportional to the number n of assets, as opposed to quadratic in n. Finally, in an numerical experiment we examine the impact of stochastic correlations on the steepness of the implied volatility smile of index options. To avoid the curse of dimensionality when pricing basket options, we propose two solutions. The first one is to use standard Monte Carlo techniques. The second one is to solve the high dimensional partial differential equation that option prices satisfy using the Quantized Tensor Train representation for large matrices entering in the Finite Difference discretization. This low parametric format for high dimensional tensors makes it possible for the storage cost and computational complexity to grow linearly with the number of assets.

KEYWORDS: Multi-asset framework, stochastic correlations, basket option pricing, high dimensional PDE, (Quantized) Tensor Train representation.

1 Introduction

Modelling the dynamics of the dependencies between assets is necessary to consistently select portfolios of assets as well as price and hedge derivatives consistently on these individual and multiple underlying assets. In particular, options on several underlyings (such as index options, best-of options, etc.) can be appealing financial derivatives for investors who wish to diversify their portfolios. However, this diversification benefit can dramatically vary depending on market factors and the state of the economy in general. In fact, there is substantial evidence that correlation between asset returns evolves randomly (see for instance Bollerslev, Engle, and Wooldridge (1988), Moskowitz (2003)) and that correlations tend to increase when returns are negative. On a worldwide scale, the financial crisis of 2008 has revealed how interrelated banks and insurances are and how correlation between companies' returns can rise due to the presence of systemic risk. In fact, the possibility of equity correlations increasing in the event of an economic downturn is thought to generate a significant premium for protective deep out of the money index put options, thus providing an explanation for the steepness of the index implied volatility smile.

Despite the importance of describing and predicting the *joint* behavior of assets, there are relatively few multi-asset models. The reasons are mainly technical. First, capturing the joint dynamics of many assets generally means adding stochastic factors, hence increasing dramatically the computational power needed for estimation or calibration. Second, it is theoretically challenging to build a model for which asset correlations evolve randomly and yet the stochastic correlation matrix remains almost surely positive definite over time.

The objective of this paper is to propose a framework which tackles both of these technical issues and to demonstrate its relevance in a financial setting. Our contribution is threefold: First, we present a stochastic correlation framework that can generally be used to introduce the dependence between diffusion processes. We show that the system of stochastic differential equations (SDEs) driving the correlations admits a unique strong solution and that the resulting correlation matrix is almost surely positive definite. We apply this framework to correlate stock returns each following a general stochastic volatility process and show that this general multi-asset framework also admits a unique strong solution. Second, we develop a Finite Difference scheme to price options on a basket of stocks - in particular on an equity index - using the recently developed Tensor Train format. This low parametric representation of high dimensional tensors makes it possible to have memory and computational costs that scale linearly in the number of dimensions for the PDE. Finally, we investigate the impact of stochastic correlations on the shape of the index implied volatility skew.

The model we propose for the joint dynamics of assets is very intuitive since the assets are correlated via common economic factors. This framework is based on the work of Kaya Boortz (2008) that we have extended to include multiple common factors and stochastic volatility for the individual stocks dynamics. Our resulting multi-asset model is very general and flexible. In particular, the dynamics of each asset need not be the same and can be specified freely provided it has a diffusion part. Furthermore, our paper is related to the literature using stochastic correlation processes such as van

Emmerich (2006), Ma (2009), Veraart and Veraart (2012). The main difference is that they model only one correlation process since they focus on the dynamics of two assets only (or two processes, e.g. returns and volatility) and do not mention how to extend it to a stochastic correlation matrix. A first generalization to a matrix correlation was developed by Driessen, Maenhout, and Vilkov (2009) where all assets are correlated using the same stochastic process. A more flexible generalization of a Jacobi type stochastic correlation process was recently developed by Ahdida and Alfonsi (2013). The model we use is different because it is directly built from univariate Jacobi processes using common factors that have a financial interpretation (such as systemic risk or risks inherent to industry sectors). Their multivariate extension is a priori not related to the presence of common factors but represents an interesting alternative to model stochastic correlations for financial applications. An important difference however is that our common factor framework is more parsimonious as it enables us to simulate paths of a random correlation matrix by simulating $p \times n$ univariate Brownian motions (where n is the number of assets and p the number of common factors) when Ahdida and Alfonsi (2013) need to simulate n^2 Brownian motions.

Another related strand of literature has developed and used stochastic covariance matrices as opposed to correlation matrices. In particular, Bru (1991) introduced SDEs defined on positive semidefinite matrices for the Wishart process that were later extended to include jumps in Leippold and Trojani (2008), Cuchiero, Filipovic, Mayerhofer, and Teichmann (2009). The Wishart process (and extensions) has a closed form characteristic function which is important for estimation and option pricing applications. To cite only a few, Da Fonseca, Grasselli, and Tebaldi (2007), Leippold and Trojani (2008) and Gourieroux, Jasiak, and Sufana (2009) have used the Wishart process to model the stochastic covariance between assets. It turns out that despite the existence of a closed form characteristic function, it is computationally too intensive to use this model to price options on many assets. The main reason is that the number of factors is $O(n^2)$. This renders models based on Wishart processes extremely slow to estimate or calibrate and additionally makes the identification of $O(n^2)$ parameters very intricate. Moreover, the choice of modelling the covariance of assets instead of correlations implies that the dependence dynamics of assets is mixed together with the univariate dynamics of individual assets. This means that the dynamics of any particular asset depends on all $O(n^2)$ stochastic factors which is neither intuitive, nor computationally efficient. As a consequence, our model involving only $O(n)$ factors represents a promising alternative to Wishart type models. Indeed, if one wishes to estimate our model, one can as a first step estimate each individual asset's dynamics independently - and in parallel - using time series or options' data. Then, in a second step, the dependence structure can be estimated via derivatives on baskets or historical correlations. As a consequence, this setup guaranties an estimation of at most $O(n)$ factors simultaneously.

The last strand of literature related to our work is the pricing of index options and the research attempting to explain the steepness of index implied volatilities compared to individual assets implied volatilities. Langnau (2010) shows that it is necessary to have stochastic correlations to consistently fit an index implied volatility smile and its components' smiles. His approach corresponds to local correlation modelling, i.e. modelling stochastic correlations as a deterministic function of the index or

a function of its components' stock prices. In the local correlation setup, the prices of basket options help identify the shape of the deterministic local correlation function. One issue with local correlation models is that the resulting stochastic correlation matrix may fail to be positive semidefinite. An extension that avoids this issue is developed by Guyon (2013). However, despite their ability to fit market data, one important drawback of local correlation models is the lack of interpretation and plausibility for the shape of the calibrated local correlation function. As in the case of the local volatility models, the local function adapts to fit the problem at hand but generates correlation patterns (as a function of the underlying stocks) that may not make sense. Another alternative to fit an index and its individual components' implied volatility smiles is proposed by Cont and Deguest (2010). They find that introducing a common jump to all assets - in a multivariate extension of the Merton model - makes it possible to fit the steep implied volatilities of index options together with the flatter smiles of individual components. In fact, rather than advocating the use of a particular model, Cont and Deguest (2010) design a calibration method to consistently price index and component options as well as quantify model uncertainty. In particular, one could estimate our model using their calibration technique.

To price index options in our framework, one can use Monte Carlo techniques which is appropriate for high dimensional problems. We present an alternative by solving numerically the partial differential equation (PDE) that option prices satisfy using numerical techniques developed in the area of tensor representations. To avoid the curse of dimensionality (i.e., the memory usage and computational complexity grows exponentially in the number of assets), we use the Tensor Train (TT) format recently introduced by Oseledets and Tyrtyshnikov (2009) and Oseledets (2011). This is a low parametric format for high dimensional tensors that it is based on robust algorithms and is free of the curse of dimensionality provided ranks of unfolding matrices are bounded as a function of the number of dimensions. All basic operations (matrix-vector multiplication, linear solver, etc.) are implemented and available publicly at `http://spring.inm.ras.ru/osel/?page_id=24`.

This rest of the paper is organized as follows. Section 2 introduces the stochastic correlation framework and a general multi-asset stochastic volatility model of stochastically correlated assets. Existence and uniqueness of a solution to the SDEs proposed are shown and important properties of this framework presented. Section 3 presents the numerical techniques used to price multi-asset options in this framework. Although we present boundary conditions for the truncated PDE that correspond to an index put, one can adapt very easily these conditions to price any multi-asset option, including path-dependent options. In this section, we present in particular the Tensor Train format which helps to avoid to some extent the curse of dimensionality when discretizing the PDE. Finally, section 4 presents numerical experiments and shows that stochastic correlation between assets generates a steeper index implied volatility smile compare to models with constant correlations.

2 Framework and properties

This section introduces a general framework to model n dependent assets. To keep the setup intuitive, we assume that the assets are stocks.

2.1 Setup and model intuition

We consider a filtered probability space $(\Omega, \mathcal{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. The measure $\mathbb{P}$ corresponds to the historical probability measure where we observe n stock prices S^i evolve over time $t \geq 0$. The stock price processes S^i are adapted to the filtration $\mathcal{F}$. Assuming there is no arbitrage in this financial market, the fundamental theorem of asset pricing guaranties the existence of a martingale measure $\mathbb{Q}$. In this setting, the martingale measure is not unique but we assume that it has been fixed[38].

The risk-neutral dynamics of the n asset-returns $Y_t := (Y_t^1, ..., Y_t^n)^\top = (\ln S_t^1, ..., \ln S_t^n)^\top$ is given by

$$dY_t = \left(r - \frac{1}{2}\text{Diag}\left(v_t^1, \cdots, v_t^n\right)\right) 1_{\mathbb{R}^n} dt + \sqrt{\text{Diag}\left(v_t^1, \cdots, v_t^n\right)} dW_t, \tag{1}$$

where r is the risk-free interest rate (assumed to be constant). The notation Diag() is a diagonal matrix where the diagonal is built from the input vector. W is a vector of n Brownian motions. We later introduce the dependence structure of the Brownian increments dW to have stochastically correlated assets. We assume that the dynamics of the instantaneous variance v^i is known and that these dynamics *only depend on asset i* in the form

$$dv_t^i = \kappa_i(\theta_i - v_t^i)dt + b_i(v_t^i)\left[\zeta^i dW_t^i + \sqrt{1 - (\zeta^i)^2} d\hat{B}_t^i\right], \tag{2}$$

where the drift is affine in v^i assuring the mean-reversion property of variance, with speed of reversion $\kappa_i > 0$ and reversion level θ_i. b_i is a function of the variance depending on the specification we wish to have[39]. $\{\hat{B}^i\}_{1\leq i\leq n}$ are independent Brownian motions independent of $\{W^i\}_{1\leq i\leq n}$. The correlation parameter $\zeta^i \in [-1, 1]$ introduces for each asset i a leverage effect. The parameters of the variance process v^i and the function b_i need to satisfy conditions to ensure positivity (and no explosion) of v^i and existence of a strong solution to this SDE. We provide details in section 2.2.

At this stage, this is a standard continuous - yet fairly general - set up. It is furthermore easily possible to incorporate jumps in the asset dynamics, however, these additional features are not necessary for our purpose to construct stochastically correlated assets and we therefore focus on the description

[38] For example, one can choose a risk-neutral measure by minimizing a specific distance between model and market option prices.

[39] For instance $b_i(x) = \nu_i\sqrt{x}$ corresponds to the Heston model and $b_i(x) = \nu_i x$ to the Hull-White model.

of the dependence structure[40].

We note that the dependence structure is hidden in W and does not impact the dynamics of individual assets. In particular, we can specify different dynamics for each asset[41].

We now introduce the dependence structure between these assets. Consider n stochastic correlation processes defined by

$$d\rho_t^i = \gamma_i(\rho_\infty^i - \rho_t^i)dt + \xi_i\sqrt{1-\left(\rho_t^i\right)^2}\left[\delta^i d\bar{W}_t^0 + \sqrt{1-(\delta^i)^2}d\hat{R}_t^i\right], \quad \forall i \in \{1,...,n\} \tag{3}$$

$$\xi_i^2 < \gamma_i\left(1 \pm \rho_\infty^i\right), \quad \forall i \in \{1,...,n\} \tag{4}$$

This process is obtained after an affine transformation of a Jacobi process[42], it therefore remains in the interval $(-1,1)$ almost surely provided $\rho_\infty^i \in (-1,1)$ and the condition (4) is satisfied. $\bar{W}^0$ is a Brownian motion that is a common factor to all correlations and the Brownian motion $\hat{R}^i$ corresponds to individual shocks in correlation i that are independent of the process $\bar{W}^0$. The parameters in (3) additionally satisfy $\gamma_i, \xi_i > 0$ and $\delta^i \in [-1,1]$. We note that the independent Brownian motions $\{\hat{R}^i\}_{1\leq i\leq n}$ are independent of $\{\hat{B}^i\}_{1\leq i\leq n}$ and $\{W^i\}_{1\leq i\leq n}$.

Furthermore, the returns of asset i have random shocks defined by

$$dW_t^i = \rho_t^i d\bar{W}_t^0 + \sqrt{1-\left(\rho_t^i\right)^2}\left[\sum_{k=1}^p \alpha_k^i d\bar{W}_t^k + \beta^i d\hat{W}_t^i\right], \tag{5}$$

where $\bar{W} := \left(\bar{W}^1, \cdots, \bar{W}^p\right)^\top$ is a vector Brownian motion with independent components that represents common factors impacting the returns of all assets. For each i, the vector $\alpha^i := \left(\alpha_1^i, \cdots, \alpha_p^i\right)^\top \in (-1,1)^p$ introduces a correlation between W^i and the common factors $\bar{W}$. Finally, the returns of asset i are impacted by firm-specific shocks represented by the Brownian motion $\hat{W}^i$ independent of all other processes. The parameters α^i and β^i have to satisfy $\sum_{k=1}^p \left(\alpha_k^i\right)^2 + \left(\beta^i\right)^2 = 1$, for each $i \in \{1,...,n\}$ with $\beta^i \neq 0$. The assumption $\beta^i \neq 0$ is important if we wish to have a full rank correlation matrix between assets. When $\rho^i, \rho^j \to \pm 1$ (for $i \neq j$) then assets are perfectly can become perfectly correlated via the common factor $\bar{W}^0$. Alternatively, if $\beta^i, \beta^j \to 0$ (for $i \neq j$) and the vectors $\alpha^i = \alpha^j$ and the processes ρ^i, ρ^j are the same, then assets i and j become perfectly correlated. The case of very high correlation between assets is important for basket option pricing but assuming that some assets can become perfectly correlated is not realistic. We therefore prevent these degenerate cases from happening by choosing that $\beta^i \neq 0$ for all $i \in \{1,...,n\}$. We note that since the variance process v^i of asset i is correlated to the returns Y^i, the stochastic variance processes $\{v^i\}_{1\leq i\leq n}$ are also indirectly

[40] Additionally, estimation of dependent stock price processes with stochastic volatility and dependent jumps raises serious issues of identifiability of the parameters (to disentangle stochastic volatility from jumps) unless estimating processes with high frequency data. We therefore restrict ourself to continuous stochastic processes.

[41] This feature is advantageous as it allows to use existing implementation of models for individual assets and introduce dependence afterwards.

[42] Jacobi processes are for instance studied in Gourieroux and Jasiak (2006), Kaya Boortz (2008) or more recently in Veraart and Veraart (2012).

stochastically correlated.

We interpret the Brownian motion $\bar{W}^0$ as a correlation random factor which represents part of the systemic risk in the pool of assets. It impacts the correlation between assets through equation (3) but also the returns in each asset (5). Indeed, we note that negative increments $d\bar{W}^0$ can increase the correlation between all assets (assuming $\delta^i < 0$ in (3)) as well as cause negative shocks in all asset dynamics (5). The other common factors $\left(\bar{W}^1, \cdots, \bar{W}^p\right)$ represent parts of sectors not affected by systemic risk to which companies may be connected to.

Remark 2.1. *Provided the number of factors $p+1$ is chosen independently of n (not scaled with n), the number of parameters is $O(n)$. An advantage of this framework is that regardless of p, the number of stochastic factors is always $O(n)$.*

This framework is an extension of the common factor model presented in Kaya Boortz (2008), where we have introduced additional common factors (the case $\alpha^i = 0, \forall i$, corresponds to Kaya Boortz (2008)) and stochastic volatility for each asset. In section 4.1, we explain why we should have $p+1 \geq 2$ (i.e., more than one common factor) when modelling $n \geq 3$ assets.

2.2 Model properties

We first show that the system of stochastic correlation processes (3) is well specified, i.e. that there exists a unique strong solution to SDEs (3). We prove this result before showing the existence of a solution to the asset SDEs since the specification of the stochastic correlations processes is independent of the asset returns processes and can therefore be used in a different setup.

Theorem 2.1. Existence of a solution for the system of SDEs (3)**.**
Assuming conditions (4) *are satisfied for each i, i.e., $\mathbb{Q}(\rho_t^i \in [-1,1], \forall t \geq 0) = 1 \ \forall i \in \{1, ..., n\}$, then there exists a unique strong solution for the system of SDEs* (3)*.*

Proof. We will apply theorem 1.1 of Luo (2011) which we recall in the Appendix 1. For this purpose, we re-write the system of SDEs (3) into one vector SDE:

$$dX_t = \eta(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = (\rho_0^1, ..., \rho_0^n)^\top \in (-1,1)^n,$$

where $X_t = (\rho_t^1, ..., \rho_t^n)^\top$, B is the vector of standard Brownian motions $B = (\bar{W}^0, \hat{R}^1, ..., \hat{R}^n)^\top$,

$\eta(x) = \left(\gamma_1(\rho^1_\infty - x_1), ..., \gamma_n(\rho^n_\infty - x_n)\right)^\top$ and σ is the matrix

$$\sigma(x) := \begin{pmatrix} \text{Vect}\left(\xi_i\sqrt{1-x_i^2}\delta^i\right) & \text{Diag}\left(\xi_i\sqrt{1-x_i^2}\sqrt{1-(\delta^i)^2}\right) \end{pmatrix}$$

for $x \in (-1,1)^n$. Vect() is a column vector of size n and Diag() a diagonal matrix of size n. As a consequence $\sigma(x) \in \mathcal{M}_{n,n+1}$.

Given that we assume (4), boundaries -1 and 1 are unattainable for every process $(\rho^i_t)_{t\geq 0}$.

We start by showing that assumption 1.1 is satisfied. Consider $x, y \in (-1,1)^n$,

$$||\sigma(x)^{i,\cdot} - \sigma(y)^{i,\cdot}||_2 = \left(\xi_i^2\left(\delta^i\right)^2\left|\sqrt{1-x_i^2} - \sqrt{1-y_i^2}\right|^2 + \xi_i^2\left(1-\left(\delta^i\right)^2\right)\left|\sqrt{1-x_i^2} - \sqrt{1-y_i^2}\right|^2\right)^{1/2},$$

where $\sigma(x)^{i,\cdot}$ denotes the i-th row of the matrix $\sigma(x)$. We use that $|\sqrt{|a|} - \sqrt{|b|}| \leq \sqrt{|a-b|}$ for any $a, b \in \mathbb{R}$ and get

$$\begin{aligned} ||\sigma(x)^{i,\cdot} - \sigma(y)^{i,\cdot}||_2 &\leq \xi_i\sqrt{|x_i^2 - y_i^2|} = \xi_i\sqrt{|x_i - y_i|}\sqrt{|x_i + y_i|} \leq \xi_i\sqrt{2}\sqrt{|x_i - y_i|} \\ &\leq \xi_i\sqrt{2}\left(\sqrt{|x_i - y_i|^2}\right)^{1/2} \leq \xi_i\sqrt{2}\left(\sum_{k=1}^{i}|x_k - y_k|^2\right)^{1/2} \end{aligned}$$

which corresponds to assumption 1.1 with $f(u) = \xi_i\sqrt{2}\sqrt{u}$. Clearly f is strictly increasing, starts at 0 and $u \mapsto \frac{1}{2\xi_i^2 u}$ is non-integrable in the neighborhood of 0^+.

The function $x \mapsto \eta(x)$ is affine and therefore Lispchitz. The assumption 1.2 is satisfied with $g = id$.

Finally, given that the functions $x \mapsto \eta(x)$ and $x \mapsto \sigma(x)$ are continuous and that $x \in (-1,1)^n$, they are therefore bounded and satisfy the linear growth condition 1.3.

As a consequence of theorem 1.1, there exists a unique strong solution to the system of SDEs (3). □

Assumption 2.1. *Continuity assumptions on b_i.*
The functions $\{b_i\}_{1\leq i\leq n}$ in (2) satisfy the following conditions:

(i) $\forall i \in \{1, ..., n\}, b_i : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *is continuous,*

(ii) $\forall i \in \{1, ..., n\}, \exists C_i, c_0^i > 0, \forall u_1, u_2 \geq 0$

$$|u_1 - u_2| \leq c_0^i \Longrightarrow |b_i(u_1) - b_i(u_2)| \leq C_i\sqrt{|u_1 - u_2|}.$$

Assumption 2.2. *Positivity of the variance processes* $\{v^i\}_{1\leq i\leq n}$.
The variance processes $\{v^i\}_{1\leq i\leq n}$ *remain positive*[43]*, i.e.* $\mathbb{Q}(v_t^i > 0, \forall t \geq 0)$ *for all* i.

We now prove the existence of a strong solution for the asset price and stochastic variance processes.

Theorem 2.2. Existence of a solution for the system of SDEs (1) **-** (2) **-** (3)**.**
Under assumptions 2.1 and 2.2 and assuming conditions (4) *are satisfied for each* i*, the system of SDEs* (1) - (2) - (3) *admits a unique strong solution.*

Proof. The system of SDEs (3) for the correlation processes are defined independently of the SDEs (1) - (2). It therefore only remains to show pathwise uniqueness (and therefore existence of a unique strong solution) for the SDEs (1) - (2). For this purpose, we will apply theorem 1.1 which holds with conditions that are localized, see remark 1.1 of Appendix 1.

We write the vector SDE corresponding to (1) - (2) - (3).

$$dX_t = \eta(X_t)dt + \sigma(X_t)dB_t, \quad X_0 \in \mathcal{D},$$

$\mathcal{D} = (-1,1)^n \times \left(\mathbb{R}_+^*\right)^n \times \mathbb{R}^n$ where $X_t = (\rho_t^1, ..., \rho_t^n, v_t^1, ..., v_t^1, Y_t^1, ..., Y_t^n)^\top$, B is the vector of standard Brownian motions $B = (\bar{W}^0, \bar{W}^1, ..., \bar{W}^p, \hat{R}^1, ..., \hat{R}^n, \hat{B}^1, ..., \hat{B}^n, \hat{W}^1, ..., \hat{W}^n)^\top$, η is defined as

$$\eta(x) = \left(\gamma_1(\rho_\infty^1 - x_1), ..., \gamma_n(\rho_\infty^n - x_n), \kappa_1(\theta_1 - x_{n+1}), ..., \kappa_n(\theta_n - x_{2n}), r - \frac{1}{2}x_{n+1}, ..., r - \frac{1}{2}x_{2n}\right)^\top$$

and $\sigma(x)$ is the matrix in $\mathcal{M}_{3n,3n+p+1}$

$$\sigma(x) := \begin{pmatrix} \text{Vect}\left(\xi_i\sqrt{1-x_i^2}\delta^i\right) & 0_{n,p} & \text{Diag}\left(\xi_i\sqrt{1-x_i^2}\sqrt{1-(\delta^i)^2}\right) & \cdots \\ \text{Vect}\left(b_i(x_{n+i})\zeta^i x_i\right) & \left(b_i(x_{n+i})\zeta^i\sqrt{1-x_i^2}\alpha_j^i\right)_{i=1:n,j=1:p} & 0_{n,n} & \cdots \\ \text{Vect}\left(\sqrt{x_{n+i}}x_i\right) & \left(\sqrt{x_{n+i}}\sqrt{1-x_i^2}\alpha_j^i\right)_{i=1:n,j=1:p} & 0_{n,n} & \cdots \end{pmatrix}$$

$$\begin{pmatrix} \cdots & 0_{n,n} & 0_{n,n} \\ \cdots & \text{Diag}\left(b_i(x_{n+i})\sqrt{1-(\zeta^i)^2}\right) & \text{Diag}\left(b_i(x_{n+i})\zeta^i\sqrt{1-x_i^2}\beta^i\right) \\ \cdots & 0_{n,n} & \text{Diag}\left(\sqrt{x_{n+i}}\sqrt{1-x_i^2}\beta^i\right) \end{pmatrix}$$

for $x \in \mathcal{D}$. Vect() is a column vector of size n and Diag() a diagonal matrix of size n.

As stated above, we already have pathwise uniqueness for the first n components of the process X. It remains to show it for the remaining components. We first focus on the variance processes, i.e. on components of X from $n+1$ to $2n$. We use the localized version of assumption 1.1 detailed in Remark

[43]For instance, if asset i follows a Heston model, where $b_i(v_t^i) = \nu_i\sqrt{v_t^i}$, the condition is $2\kappa_i\theta_i > \nu_i^2$.

1.1. Consider $x, y \in \mathcal{D}$ and $i \in \{1, ..., n\}$. We have

$$\begin{aligned}
\left|\left|\sigma^{n+i,\cdot}(x) - \sigma^{n+i,\cdot}(y)\right|\right|_2 &= \Bigg[\left(\zeta^i\right)^2 |b_i(x_{n+i})x_i - b_i(y_{n+i})y_i|^2 + \left(\zeta^i\right)^2 \left| b_i(x_{n+i})\sqrt{1-x_i^2} - b_i(y_{n+i})\sqrt{1-y_i^2} \right|^2 \\
&\quad + \left(1 - \left(\zeta^i\right)^2\right) |b_i(x_{n+i}) - b_i(y_{n+i})|^2 \Bigg]^{1/2} \\
&= \Big[\left(\zeta^i\right)^2 |b_i(x_{n+i})x_i - b_i(y_{n+i})x_i + b_i(y_{n+i})x_i - b_i(y_{n+i})y_i|^2 \\
&\quad + \left(\zeta^i\right)^2 \left| b_i(x_{n+i})\sqrt{1-x_i^2} - b_i(y_{n+i})\sqrt{1-x_i^2} + b_i(y_{n+i})\sqrt{1-x_i^2} - b_i(y_{n+i})\sqrt{1-y_i^2} \right|^2 \\
&\quad + \left(1 - \left(\zeta^i\right)^2\right) |b_i(x_{n+i}) - b_i(y_{n+i})|^2 \Big]^{1/2}
\end{aligned}$$

Let us consider $R > 0$. Given that the functions $\{b_i\}_{1 \le i \le n}$ are continuous, they are bounded on the compact set $[0, R]$ and we denote by $M_{b_i}(R) := \max_{[0,R]} b_i$. We now apply the triangle inequality to the terms above:

$$\begin{aligned}
\left|\left|\sigma^{n+i,\cdot}(x) - \sigma^{n+i,\cdot}(y)\right|\right|_2 &\le \Bigg[\left(\zeta^i\right)^2 \left(\underbrace{|x_i|}_{\le 1} |b_i(x_{n+i}) - b_i(y_{n+i})| + \underbrace{|b_i(y_{n+i})|}_{\le M_{b_i}(R)} |x_i - y_i| \right)^2 \\
&\quad + \left(\zeta^i\right)^2 \left(\underbrace{\sqrt{1-x_i^2}}_{\le 1} |b_i(x_{n+i}) - b_i(y_{n+i})| + \underbrace{|b_i(y_{n+i})|}_{\le M_{b_i}(R)} \left| \sqrt{1-x_i^2} - \sqrt{1-y_i^2} \right| \right)^2 \\
&\quad + \left(1 - \left(\zeta^i\right)^2\right) |b_i(x_{n+i}) - b_i(y_{n+i})|^2 \Bigg]^{1/2}
\end{aligned}$$

Furthermore simplifying the second squared bracket

$$\left| \sqrt{1-x_i^2} - \sqrt{1-y_i^2} \right| \le \sqrt{\left|x_i^2 - y_i^2\right|} = \sqrt{|x_i - y_i||x_i + y_i|} \le \sqrt{2}\sqrt{|x_i - y_i|}.$$

Each square bracket can be bounded from above using $(a+b)^2 \le 2(a^2+b^2)$ and then using that $|x_i - y_i|^2 \le 2|x_i - y_i|$ and regrouping the terms we get

$$\left|\left|\sigma^{n+i,\cdot}(x) - \sigma^{n+i,\cdot}(y)\right|\right|_2 \le \left[\left(3\left(\zeta^i\right)^2 + 1\right) |b_i(x_{n+i}) - b_i(y_{n+i})|^2 + 8\left(\zeta^i\right)^2 M_{b_i}(R)^2 |x_i - y_i| \right]^{1/2}.$$

We now use the second part of assumption 2.1. For each $i \in \{1, ..., n\}$ there exists a constant $c_0^i > 0$ such that $|u_1 - u_2| \le c_0^i \Longrightarrow |b_i(u_1) - b_i(u_2)| \le C_i\sqrt{|u_1 - u_2|}$ for some constant $C_i > 0$ independent

of R. We define $c_0 := \min_{1\leq i\leq n}\{c_o^i\}$. For $||x-y||_2 \leq c_0$, we therefore have

$$\begin{aligned}\left|\left|\sigma^{n+i,\cdot}(x)-\sigma^{n+i,\cdot}(y)\right|\right|_2 &\leq \left[\left(3\left(\zeta^i\right)^2+1\right)C_i^2\left|x_{n+i}-y_{n+i}\right|+8\left(\zeta^i\right)^2 M_{b_i}(R)^2|x_i-y_i|\right]^{1/2}\\ &\leq \left(\max\left(\left(3\left(\zeta^i\right)^2+1\right)C_i^2, 8\left(\zeta^i\right)^2 M_{b_i}(R)^2\right)\right)^{1/2}\left[\sum_{k=1}^{n+i}|x_k-y_k|\right]^{1/2}\end{aligned}$$

Given that all norms are equivalent in $\mathbb{R}^{n+i}$, we get

$$\left|\left|\sigma^{n+i,\cdot}(x)-\sigma^{n+i,\cdot}(y)\right|\right|_2 \leq C_R^i\left[\left(\sum_{k=1}^{n+i}|x_k-y_k|^2\right)^{1/2}\right]^{1/2} = C_R^i f\left(\sqrt{\sum_{k=1}^{n+i}|x_k-y_k|^2}\right)$$

where $C_R^i = \left(\max\left(\left(3\left(\zeta^i\right)^2+1\right)C_i^2, 8\left(\zeta^i\right)^2 M_{b_i}(R)^2\right)\right)^{1/2}$ and the function f satisfies the assumptions of Remark 1.1.

To show that $\left|\left|\sigma^{2n+i,\cdot}(x)-\sigma^{2n+i,\cdot}(y)\right|\right|_2$ satisfies a similar inequality, we simply note that this difference is equal to $\left|\left|\sigma^{n+i,\cdot}(x)-\sigma^{n+i,\cdot}(y)\right|\right|_2$ for $\zeta^i = 0$ and $b_i(u) = \sqrt{u}$. Hence we can find a similar inequality.

Furthermore, assumption 1.2 is clearly satisfied given that the each component $\eta_i(x)$ is affine in one component of x and hence Lipschitz.

It remains to show that η and σ satisfy the linear growth condition of assumption 1.3. $\eta(x)$ is affine in the components of x and therefore satisfies the condition. For σ, it is sufficient to show it for $\sigma(x)^{n+i,\cdot}$, with $i \in \{1,...,n\}$. Simple calculations lead to

$$\left|\left|\sigma^{n+i,\cdot}(x)\right|\right|_2^2 = \left|b_i(x_{n+i})\right|^2.$$

We now show that $\left|b_i(x_{n+i})\right|^2$ satisfies the linear growth condition with respect to the variable x_{n+i}. Let us denote by N_x the integer part of x_{n+i}/c_0^i, we have

$$|b_i(x_{n+i})| \leq |b_i(0)| + \sum_{k=0}^{N_x-1}\underbrace{\left|b_i\left((k+1)c_0^i\right)-b_i\left(kc_0^i\right)\right|}_{\leq C_i\sqrt{\left|(k+1)c_0^i-kc_0^i\right|}} + \underbrace{\left|b_i\left(x_{n+i}\right)-b_i\left(N_x c_0^i\right)\right|}_{\leq C_i\sqrt{\left|x_{n+i}-N_x c_0^i\right|}}$$

where the sum is 0 if $N_x \leq 1$.

$$|b_i(x_{n+i})| < |b_i(0)| + C_i N_x\sqrt{c_o^i} + C_i c_0 = \left(|b_i(0)| + C_i c_0\right) + \frac{C_i}{\sqrt{c_o^i}}x_{n+i}$$

which shows that the linear growth condition of $\left|\left|\sigma^{n+i,\cdot}(x)\right|\right|_2$ is satisfied or equivalently the quadratic growth of $\left|\left|\sigma^{n+i,\cdot}(x)\right|\right|_2^2$. This concludes the proof. □

Proposition 2.1. *For every $i \in \{1, ..., n\}$, W^i defined in* (5) *is a $\{\mathcal{F}_t\}_{t\geq 0}$ adapted Brownian motion.*

Proof. Let us fix one $i \in \{1, ..., n\}$ and denote by $\tilde{W}$ the Brownian motion $\tilde{W}_t := \left[\sum_{k=1}^p \alpha_k^i d\bar{W}_t^k + \beta^i d\hat{W}_t^i\right]$. We can write

$$dW_t^i = \rho_t^i d\bar{W}_t^0 + \sqrt{1 - \left(\rho_t^i\right)^2} d\tilde{W}_t,$$

where $\bar{W}^0$ and $\tilde{W}$ are independent Brownian motions.

By definition, the process ρ^i is $\mathcal{F}$ adapted and bounded in $[-1, 1]$. Additionally, since $\bar{W}^0$ and $\tilde{W}$ are two independent Brownian motions, the stochastic integrals $\int_0^t \rho_s^i d\bar{W}_s^0$ and $\int_0^t \sqrt{1 - (\rho_s^i)^2} d\tilde{W}_s$ are continuous independent martingales.

We now show that $\mathbb{E}\left[(W_t^i)^2\right] = t$. Using Itô's isometry, we have

$$\begin{aligned}
\mathbb{E}\left[(W_t^i)^2\right] &= \mathbb{E}\left[\left(\int_0^t \rho_s^i d\bar{W}_s^0\right)^2 + \left(\int_0^t \sqrt{1 - (\rho_s^i)^2} d\tilde{W}_s\right)^2 + 2\int_0^t \rho_s^i d\bar{W}_s^0 \int_0^t \sqrt{1 - (\rho_s^i)^2} d\tilde{W}_s\right] \\
&= \mathbb{E}\left[\int_0^t \left(\rho_s^i\right)^2 ds + \int_0^t 1 - \left(\rho_s^i\right)^2 ds\right] + 2\underbrace{\mathbb{E}\left[\int_0^t \rho_s^i d\bar{W}_s^0\right]}_{=0}\underbrace{\mathbb{E}\left[\int_0^t \sqrt{1 - (\rho_s^i)^2} d\tilde{W}_s\right]}_{=0} \\
&= t.
\end{aligned}$$

Since $W_0^i = 0$, we can conclude that W^i is a Brownian motion. □

Proposition 2.2. *The model* (1) - (5) *has the following properties*

(i) $d\left[Y^i, Y^i\right]_t = v_t^i dt \quad , \quad d\left[v^i, v^i\right]_t = \left(b_i(v_t^i)\right)^2 \quad , \quad d\left[\rho^i, \rho^i\right]_t = \xi_i^2 \left(1 - \left(\rho_t^i\right)^2\right) dt,$

(ii) $d\left[W^i, W^j\right]_t = \left(\rho_t^i \rho_t^j + \sqrt{1 - \left(\rho_t^i\right)^2}\sqrt{1 - \left(\rho_t^j\right)^2} \sum_{k=1}^p \alpha_k^i \alpha_k^j\right) dt, \quad i \neq j.$

(iii) *For $i \neq j$, we have*

$$\begin{aligned}
d\left[Y^i, Y^j\right]_t &= \sqrt{v_t^i}\sqrt{v_t^j} d\left[W^i, W^j\right]_t, \\
d\left[v^i, v^j\right]_t &= b_i(v_t^i) b_j(v_t^j) \zeta^i \zeta^j d\left[W^i, W^j\right]_t.
\end{aligned}$$

In particular, the instantaneous correlation between (Y^i, Y^j) *is*

$$C_t^{ij} := \frac{d\left[Y^i, Y^j\right]_t}{\sqrt{d\left[Y^i, Y^i\right]_t d\left[Y^j, Y^j\right]_t}} = \frac{d\left[W^i, W^j\right]_t}{dt} = \rho_t^i \rho_t^j + \sqrt{1-\left(\rho_t^i\right)^2}\sqrt{1-\left(\rho_t^j\right)^2}\sum_{k=1}^{p}\alpha_k^i\alpha_k^j\ , \quad i \neq j. \tag{6}$$

Proposition 2.3. *The correlation matrix process* C_t *for the Brownian motions* $(W^1, \cdots, W^n)$ *defined by the model* (1) - (3) - (5) *is continuous and almost surely positive definite on any interval* $[0, T]$*, for any* $T > 0$*.*

Proof. Using proposition 2.2 we know that the matrix process C is defined by

$$C_t^{ii} = 1, \quad \forall i \in \{1, \cdots, n\}, C_t^{ij} = \rho_t^i \rho_t^j + \sqrt{1-\left(\rho_t^i\right)^2}\sqrt{1-\left(\rho_t^j\right)^2}\sum_{k=1}^{p}\alpha_k^i\alpha_k^j, \quad \forall i, j \in \{1, \cdots, n\}, i \neq j.$$

First of all, the matrix C_t is symmetric and the process $(C_t)_{t\geq 0}$ is continuous.

We recall the notation $\alpha^i = \left(\alpha_1^i, \cdots, \alpha_p^i\right)^\top$. For $i \neq j$, we then have

$$\begin{aligned}
|C_t^{ij}| &\leq |\rho_t^i \rho_t^j| + \sqrt{1-\left(\rho_t^i\right)^2}\sqrt{1-\left(\rho_t^j\right)^2}\underbrace{\left|\sum_{k=1}^{p}\alpha_k^i\alpha_k^j\right|}_{=|(\alpha^i)^\top(\alpha^j)|} \\
&\leq |\rho_t^i \rho_t^j| + \sqrt{1-\left(\rho_t^i\right)^2}\sqrt{1-\left(\rho_t^j\right)^2}\sqrt{(\alpha^i)^\top(\alpha^i)}\sqrt{(\alpha^j)^\top(\alpha^j)} \quad \text{(Cauchy-Schwartz's inequality)} \\
&\leq |\rho_t^i \rho_t^j| + \sqrt{1-\left(\rho_t^i\right)^2}\sqrt{1-\left(\rho_t^j\right)^2} \quad \text{using} \quad \sum_{k=1}^{p}\left(\alpha_l^k\right)^2 + \left(\beta^l\right)^2 = 1
\end{aligned}$$

Since $\rho_t^i \in [-1, 1]$ for all i and all t, we can find $\theta_t^i, \theta_t^j \in [0, \pi/2]$ such that $|\rho_t^i| = \cos\left(\theta_t^i\right)$ and $|\rho_t^j| = \cos\left(\theta_t^j\right)$ which leads to

$$|C_t^{ij}| \leq \cos\left(\theta_t^i\right)\cos\left(\theta_t^j\right) + \sin\left(\theta_t^i\right)\sin\left(\theta_t^j\right) = \cos\left(\theta_t^i - \theta_t^j\right) \in [0, 1].$$

This implies that all non-diagonal coefficients are in $[-1, 1]$.

We now need to show that the matrix C_t is almost surely positive definite. We denote by $\alpha \in \mathbb{R}^{p\times n}$ the $p \times n$ real matrix with columns $\{\alpha^i\}_{i=1,\ldots,n}$.

We have

$$C_t = \frac{d\left[W, W^\top\right]_t}{dt} = \rho_t \rho_t^\top + \text{Diag}\left(\sqrt{1-\left(\rho_t^i\right)^2}\right) \times \left[\alpha^\top \alpha + \text{Diag}\left(\left(\beta^i\right)^2\right)\right] \text{Diag}\left(\sqrt{1-\left(\rho_t^i\right)^2}\right)$$
$$= \rho_t \rho_t^\top + \text{Diag}\left(\sqrt{1-\left(\rho_t^i\right)^2}\right) \alpha^\top \alpha \text{Diag}\left(\sqrt{1-\left(\rho_t^i\right)^2}\right) + \text{Diag}\left(\left(\beta^i\right)^2\left(1-\left(\rho_t^i\right)^2\right)\right).$$

C_t is therefore the sum of three positive semi-definite matrices and hence positive semi-definite. Moreover, the dynamics (3) of the correlation processes ρ^i guaranty that the processes ρ^i almost surely never reach their boundaries ± 1. Since we assumed that $\beta^i \neq 0$, the matrix $\text{Diag}\left(\left(\beta^i\right)^2\left(1-\left(\rho_t^i\right)^2\right)\right)$ is almost surely positive definite. Therefore C_t is almost surely positive definite. □

Remark 2.2. *From the proof of Proposition 2.3, we note that a sufficient condition for this model to have a full rank correlation matrix at any time is to assume that the firm specific noise $\hat{W}$ almost surely contribute to the shocks in the returns (i.e. $\beta^i \neq 0$ and $\mathbb{Q}\left(\rho_t^i \neq \pm 1, \forall t \leq T\right) = 1$ for any $T \geq 0$). If this were not the case, we would need $p+1 \geq n$ common factors $(\bar{W}^0, \cdots, \bar{W}^p)$ to guaranty that the correlation matrix between returns is invertible.*

2.3 Multi-asset option pricing

We are interested in the pricing and risk management of a European contingent claim on the asset $S^1, ..., S^n$ with dynamics (1). In this section, we derive the infinitesimal generator of the vector process $(S^1, ..., S^n)$ and the partial differential equation that a multivariate contingent claim must satisfy in this framework.

Proposition 2.4. *The infinitesimal generator L^X of the process $X = (Y^1, v_1, \rho_1, ..., Y^n, v_n, \rho_n)$ is given by*

$$L^X g(x) = \sum_{i=1}^n \left[\frac{\partial g}{\partial y_i}(r - \frac{1}{2}v_i) + \frac{\partial g}{\partial v_i}\kappa_i(\theta_i - v_i) + \frac{\partial g}{\partial \rho_i}\gamma_i(\rho_\infty^i - \rho_i)\right] \tag{7}$$
$$+\frac{1}{2}\sum_{i=1}^n \left[\frac{\partial^2 g}{\partial y_i^2}v_i + \frac{\partial^2 g}{\partial v_i^2}b_i(v_i)^2 + \frac{\partial^2 g}{\partial \rho_i^2}\xi_i^2(1-\rho_i^2)\right]$$
$$+\sum_{i,j=1}^n \left[\frac{\partial^2 g}{\partial y_i \partial \rho_j}\sqrt{v_i}\rho_i\xi_j\sqrt{1-\rho_j^2}\delta_j + \frac{\partial^2 g}{\partial y_i \partial v_j}\sqrt{v_i}b_j(v_j)\zeta_j C_{ij}(\rho_i,\rho_j) + \frac{\partial^2 g}{\partial v_i \partial \rho_j}b_i(v_i)\zeta_i\rho_i\xi_j\sqrt{1-\rho_j^2}\delta_j\right]$$
$$+\sum_{i<j}^n \left[\frac{\partial^2 g}{\partial y_i y_j}\sqrt{v_i}\sqrt{v_j}C_{ij}(\rho_i,\rho_j) + \frac{\partial^2 g}{\partial v_i v_j}b_i(v_i)b_j(v_j)\zeta_i\zeta_j C_{ij}(\rho_i,\rho_j) + \frac{\partial^2 g}{\partial \rho_i \rho_j}\xi_i\xi_j\sqrt{(1-\rho_i^2)(1-\rho_j^2)}\delta_i\delta_j\right]$$

where $C_{ij}(\rho_i,\rho_j) = \rho_i\rho_j + \sqrt{(1-\rho_i^2)(1-\rho_j^2)}\left(\alpha^i\right)^\top \alpha^j$ for $i \neq j$ and $C_{ii} = 1$. The variable $x = (y_1, v_1, \rho_1, ..., y_n, v_n, \rho_n) = (x_1, ..., x_d)$ belongs to a suitable domain denoted by $\mathcal{D}^X \subseteq \{\mathbb{R} \times \mathbb{R}_+ \times$

$[-1,1]\}^n$. *We can also formulate the infinitesimal generator using matrices:*

$$L^X g(x) = \frac{1}{2} tr\left[Q(x) D^2 g(x)\right] + \mu(x)^\top \nabla g(x), \tag{8}$$

where Q *is a matrix depending on* x, $D^2 = \left(\frac{\partial}{\partial x_i \partial x_j}\right)_{1\le i,j\le d}$ *is the Hessian,* $\mu(x) = (\mu_1(x), ..., \mu_d(x))^\top$ *is a column vector depending on* x, $\nabla = \left(\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_d}\right)^\top$ *the gradient, and* tr *the trace operator for square matrices. More specifically, we can write*

$$\mu_{3i-2}(x) = r - \frac{1}{2} v_i = r - \frac{1}{2} x_{3i-1} \qquad , i = 1, ..., n, \tag{9}$$

$$\mu_{3i-1}(x) = \kappa_i(\theta_i - v_i) = \kappa_i(\theta_i - x_{3i-1}) \ , i = 1, ..., n, \tag{10}$$

$$\mu_{3i}(x) \quad = \gamma_i(\rho^i_\infty - \rho_i) = \gamma_i(\rho^i_\infty - x_{3i}) \ , i = 1, ..., n, \tag{11}$$

and Q *is the matrix symmetric function from* $\mathcal{D}^X$ *to* $\mathbb{R}^{d\times d}$ *defined by*

$$Q(x) := \begin{pmatrix} \tilde{Q}_{1,1}(x) & \tilde{Q}_{1,2}(x) & \cdots \\ \tilde{Q}_{2,1}(x) & \ddots & \\ \vdots & & \tilde{Q}_{n,n}(x) \end{pmatrix} \tag{12}$$

where $\tilde{Q}_{ii}$ *is a symmetric matrix function from* $\mathcal{D}^X$ *to* $\mathbb{R}^{3\times 3}$ *defined for* $i = 1, ..., n$ *by*

$$\begin{aligned} \tilde{Q}_{ii} &:= \begin{pmatrix} Q_{3i-2,3i-2} & Q_{3i-2,3i-1} & Q_{3i-2,3i} \\ Q_{3i-1,3i-2} & Q_{3i-1,3i-1} & Q_{3i-1,3i} \\ Q_{3i,3i-2} & Q_{3i,3i-1} & Q_{3i,3i} \end{pmatrix} \\ &= \begin{pmatrix} x_{3i-1} & \cdots & \cdots \\ \sqrt{x_{3i-1}} b_i(x_{3i-1})\zeta_i & b_i(x_{3i-1})^2 & \cdots \\ \sqrt{x_{3i-1}} x_{3i}\xi_i\sqrt{1-x_{3i}^2}\delta_i & b_i(x_{3i-1})\zeta_i x_{3i}\xi_i\sqrt{1-x_{3i}^2}\delta_i & \xi_i^2(1-x_{3i}^2) \end{pmatrix} \end{aligned} \tag{13}$$

and $\tilde{Q}_{ij}$ *is a matrix function from* $\mathcal{D}^X$ *to* $\mathbb{R}^{3\times 3}$ *defined for* $i, j = 1, ..., n$ *with* $i \neq j$ *by*

$$\begin{aligned} \tilde{Q}_{ij} &:= \begin{pmatrix} Q_{3i-2,3j-2} & Q_{3i-2,3j-1} & Q_{3i-2,3j} \\ Q_{3i-1,3j-2} & Q_{3i-1,3j-1} & Q_{3i-1,3j} \\ Q_{3i,3j-2} & Q_{3i,3j-1} & Q_{3i,3j} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{x_{3j-1}}\sqrt{x_{3i-1}} C_{ji}(x_{3j}, x_{3i}) & \sqrt{x_{3i-1}} b_j(x_{3j-1})\zeta_j C_{ij}(x_{3i}, x_{3j}) & \sqrt{x_{3i-1}} x_{3i}\xi_j\sqrt{1-x_{3j}^2}\delta_j \\ \sqrt{x_{3j-1}} b_i(x_{3i-1})\zeta_i C_{ji}(x_{3j}, x_{3i}) & b_i(x_{3i-1}) b_j(x_{3j-1})\zeta_i\zeta_j C_{ij}(x_{3i}, x_{3j}) & b_i(x_{3i-1})\zeta_i x_{3i}\xi_j\sqrt{1-x_{3j}^2}\delta_j \\ \sqrt{x_{3j-1}} x_{3j}\xi_i\sqrt{1-x_{3i}^2}\delta_i & b_j(x_{3j-1})\zeta_j x_{3j}\xi_i\sqrt{1-x_{3i}^2}\delta_i & \xi_i\xi_j\sqrt{(1-x_{3i}^2)(1-x_{3j}^2)}\delta_i\delta_j \end{pmatrix} \end{aligned} \tag{14}$$

Proof. We apply Itô's formula. The generator L^X is defined by the property that $\left(g(X_t) - \int_0^t L^X g(X_s) ds\right)_{t\ge 0}$

is a martingale for all g in the domain of L^X (see Ethier and Kurtz (1986)). □

Proposition 2.5. *Consider a European style contingent claim with payoff $H(X_T)$ at maturity $T > 0$. Its price h_t at time $t \geq T$ is given by the conditional expectation*

$$h_t = \mathbb{E}\left[e^{-r(T-t)}H(X_T) \mid \mathcal{F}_t\right], \tag{15}$$

where we can write the process $(h_t)_{t\geq 0}$ at time t as a function of t and X_t : $h_t = h(t, X_t)$ thanks to the Markovian property.

Provided that the function $(t, x) \mapsto h(t, x)$ belongs to the domain of the infinitesimal generator L^X, the function h is described by the following PDE:

$$\frac{\partial h}{\partial t}(t,x) + L^X h(t,x) - rh(t,x) = 0, \qquad \text{on } [0,T) \times \mathcal{D}^X \tag{16}$$

$$h(T,x) = H(x) \qquad \text{on } \mathcal{D}^X. \tag{17}$$

3 Numerical approximation of the multi-asset option price

To approximate the price h of the multivariate contingent claim[44], we propose two different techniques. The first one consists in approximating the expectation (15) by simulating the stochastic differential equations (1) - (2) - (3) - (5). For efficient simulation techniques for the Heston model, we refer to van Haastrecht and Pelsser (2010), Alfonsi (2010) and for the simulation of Jacobi type processes, we refer to Kaya Boortz (2008). The second approach, that we detail here, consists in approximating the solution to the PDE (16). Having the option price on a grid is appealing because it directly gives some of the options' greeks such as the delta or gamma. However, using standard Finite Difference (FD) techniques by discretizing each variable on an equidistant grid naturally leads to matrices whose size grows exponentially in the number of assets (the so-called curse of dimensionality). To circumvent the computational complexity, we use a recently developed low-parametric format to represent high dimensional tensors called The Tensor Train (TT) decomposition. There is a publicly available Matlab TT-toolbox (`http://spring.inm.ras.ru/osel/?page_id=24`) where basic operations on tensors (using the TT-format) as well as more developed routines solving high dimensional PDEs, such as the Poisson equation, have been implemented. In this section, we detail the discretization of the PDE (16) using Finite Differences (FD) and explain in section 3.2.2 how the resulting tensors can be approximated efficiently using the Tensor Train (TT) decomposition. For more details on this format we refer to Oseledets and Tyrtyshnikov (2009), Oseledets (2011).

[44]Since the dependence structure of the assets is specified independently of the individual assets' dynamics, it is very simple to price single asset contingent claims. If one specifies for instance Heston dynamics for each asset, one could use Fourier transform pricing for each asset and then price multivariate contingent claims using the PDE discretization.

We start by changing variables in (16) to solve the PDE forward in time and work with time to maturity $\tau = T - t$ (for $\tau \in [0, T]$) instead of t. We now want to find a function $\tilde{h} \in C^{1,2}(\mathbb{R}_+, \mathcal{D}^X)$ such that

$$\frac{\partial \tilde{h}}{\partial \tau}(\tau, x) - \mathcal{A}[\tilde{h}](\tau, x) = 0, \quad \text{on } (0, T] \times \mathcal{D}^X \tag{18}$$

$$\tilde{h}(0, x) = H(x) \quad \text{on } \mathcal{D}^X, \tag{19}$$

where the operator $\mathcal{A}$ is defined by $\mathcal{A}[\tilde{h}](\tau, x) := L^X[h](T - t, x) - rh(T - t, x)$.

To simplify notation we will write $h(\tau, x)$ instead of $\tilde{h}(\tau, x)$ in the following.

3.1 Localization and semi-discretization of the PDE

We now need to make further assumptions on the specific type of derivative being priced, i.e. on the payoff function H in Proposition 2.5, so that we can choose appropriate boundary conditions for the discretized PDE. We choose to illustrate the discretization in the case of a European put on an index, i.e. we can write the payoff $H(X_T)$ as

$$H(X_T) := \max(K - \sum_{i=1}^{n} \omega_i e^{Y_T^i}, 0) = (K - \sum_{i=1}^{n} \omega_i e^{Y_T^i})^+, \tag{20}$$

where ω_i is the weight of stock i in the index. These weights are positive and satisfy $\sum_{i=1}^{n} \omega_i = 1$. K is the strike of the option and $T > 0$ the maturity. Other choices of payoff are possible and one should adapt the artificial boundary conditions in section 3.1.1 to the corresponding payoff. Everything else remains the same.

3.1.1 Localization and artificial boundary conditions

To find a function h defined on $[0, T] \times \mathcal{D}^X$ satisfying (18), we first truncate the domain on which to find the solution.

The support of the returns y_i of stock i is $\mathbb{R}$. We denote by Ω_i^Y the truncated domain for the variable y_i. In practice, we will take $\Omega_i^Y = [-R_i, R_i]$, for some chosen $R_i > 0$.

The truncated domain of the variable v_i is denoted by $\Omega_i^v = [0, v_{max}^i]$.

The variable ρ_i $(i = 1, ..., n)$ is restricted to be in the interval $[-1, 1]$. To have consistent notation, we will denote by $\Omega_i^\rho = [-1, 1]$ the computational domain for the variables ρ_i. We do not truncate the support of the correlation processes since we are also interested in the behavior of option prices as assets become highly correlated.

We have now truncated the support of the function h satisfying (18) from the set $[0,T] \times \mathcal{D}^X$ to the truncation domain $[0,T] \times \Omega_X$, where we define $\Omega_X := \prod_{i=1}^n \{\Omega_i^Y \times \Omega_i^v \times [-1,1]\}$. To solve this new problem, we need to impose artificial boundary conditions. To ease notation in the following, we write the variable x in the generic form $x = (x_1, ..., x_d) = (y_1, u_1, \rho_1, ..., y_n, u_n, \rho_n)$, where $d = 3n$.

- The boundaries at $x_{3i-2} = y_i = R_i$, i.e. when the stock price of asset i goes to infinity, are assumed to be homogeneous Dirichlet because the option is a put. Hence we have

$$h(\tau, x_1, ..., x_{3i-2} = R_i, ..., x_d) = 0,$$

  whenever $x_{3i-2} = R_i$ for at least one $i = 1, ..., d$ and for all $\tau \in [0,T]$.

  When one stock price goes to zero, i.e. $x_{3i-2} = y_i = -R_i$, we assume a homogeneous von Neumann boundary condition, i.e.

$$\frac{\partial h}{\partial x_{3i-2}}(\tau, x_1, ..., x_{3i-2} = -R_i, ..., x_d) = 0.$$

  This is a standard boundary condition when dealing with put options, see for instance Duffy (2006).

- When the variance v_i of asset i goes to zero, i.e., $x_{3i-1} = 0$, we simply require that the discretized PDE (18) is satisfied with the operator $\mathcal{A}$ taken a the boundary point:

$$\mathcal{A}[h](\cdot) = \mathcal{A}[h](..., y_{i-1}, v_{i-1}, \rho_{i-1}, y_i, v_i = 0, \rho_i, ...),$$

  for all $y_k \in \Omega_k^Y$ $(k = 1, ..., n)$, $v_k \in \Omega_k^v$ $(k = 1, ..., n; k \neq i)$, $\rho_k \in \Omega_k^\rho$ $(k = 1, ..., n)$. A discussion of this type of boundary condition in the case of the Heston model can be found in Duffy (2006), section 22.4.

  When the variance is maximal, i.e. $x_{3i-1} = v_i = v_{max}^i$, we impose homogeneous von Neumann boundary conditions, i.e.

$$\frac{\partial h}{\partial x_{3i-1}}(\tau, x_1, ..., x_{3i-1} = v_{max}^i, ..., x_d) = 0,$$

  as in Duffy (2006).

- When the correlation $\rho_i = \pm 1$, we require that the PDE (18) is satisfied with the operator $\mathcal{A}$ taken a the boundary point:

$$\mathcal{A}^X[h](\cdot) = \mathcal{A}^X[h](..., y_{i-1}, v_{i-1}, \rho_{i-1}, y_i, v_i, \rho_i = \pm 1, ...),$$

  for all $y_k \in \Omega_k^Y$ $(k = 1, ..., n)$, $v_k \in \Omega_k^v$ $(k = 1, ..., n)$, $\rho_k \in \Omega_k^\rho$ $(k = 1, ..., n; k \neq i)$.

3.1.2 Spatial discretization

Here, we introduce a Finite Difference (FD) discretization for the truncated problem.

We start by defining grid points for the space variable $x = (y_1, u_1, \rho_1, ..., y_n, u_n, \rho_n) \in \Omega_X$, where we want to compute an approximation to the function h. For each, $i \in \{1, ..., n\}$, we define

- the grid points $\{x_{3i-2}(k)\}_{k=0,...,N_{3i-2}+1} = \{y_i(k)\} = \{-R_i + k\Delta_{3i-2}\ ,\ k = 0, ..., N_{3i-2}+1\}$ in Ω_i^Y. $\Delta_{3i-2} > 0$ is the distance between each grid point and $N_{3i-2}+2$ is the number of grid points in Ω_i^Y. By convention, we have $-R_i+(N_{3i-2}+1)\Delta_{3i-2} = R_i$ and therefore $\Delta_{3i-2} = 2R_i/(N_{3i-2}+1)$.
- the grid points $\{x_{3i-1}(k)\}_{k=0,...,N_{3i-1}+1} = \{v_i(k)\} = \{k\Delta_{3i-1}\ ,\ k = 0, ..., N_{3i-1}+1\}$ in Ω_i^v. Similarly, $\Delta_{3i-1} > 0$ is the distance between each grid point and $N_{3i-1}+2$ is the number of grid points in Ω_i^v. Again, we have $(N_{3i-1}+1)\Delta_{3i-1} = v_{max}^i$ and $\Delta_{3i-1} = v_{max}^i/(N_{3i-1}+1)$.
- the grid points $\{x_{3i}(k)\}_{k=0,...,N_{3i}+1} = \{\rho_i(k)\} = \{-1 + k\Delta_{3i}\ ,\ k = 0, ..., N_{3i}+1\}$ in Ω_i^ρ. $\Delta_{3i} > 0$ is the distance between each grid point and $N_{3i}+2$ is the number of grid points in Ω_i^ρ. We have $-1 + (N_{3i}+1)\Delta_{3i} = 1$ and $\Delta_{3i} = 2/(N_{3i}+1)$.

Our goal is to find an approximation for the solution h to (18) at the space mesh points introduced above. We now introduce a more generic notation to simplify the use of indices in the following. We define the multi-index $j = (j_1, ..., j_d)$, where $d = 3n$ for $j_l = 0, ..., N_l + 1$. We denote by h_j^τ or equivalently $h_{j_1,...,j_d}^\tau$ the approximation

$$h_j^\tau \approx h(\tau, x_1(j_1), ..., x_d(j_d)), \quad \text{for } \tau \in [0, T].$$

The second order FD discretization of the differential operators in (18) and (7) are then given by

$$\frac{\partial h}{\partial x_l}(\tau, x_1(j_1), ..., x_d(j_d)) \approx \frac{h^\tau_{j_1,...,(j_l+1),...,j_d} - h^\tau_{j_1,...,(j_l-1),...,j_d}}{2\Delta_l}, \tag{21}$$

$$\frac{\partial^2 h}{\partial x_l^2}(\tau, x_1(j_1), ..., x_d(j_d)) \approx \frac{h^\tau_{j_1,...,(j_l+1),...,j_d} - 2h^\tau_{j_1,...,j_l,...,j_d} + h^\tau_{j_1,...,(j_l-1),...,j_d}}{\Delta_l^2}, \tag{22}$$

$$\begin{aligned}\frac{\partial^2 h}{\partial x_l \partial x_i}(\tau, x_1(j_1), ..., x_d(j_d)) \approx & \frac{\left(h^\tau_{j_1,...,(j_l+1),...,(j_i+1)...,j_d} + h^\tau_{j_1,...,(j_l-1),...,(j_i-1),...,j_d}\right)}{4\Delta_l\Delta_i}, \\ & - \frac{\left(h^\tau_{j_1,...,(j_l+1),...,(j_i-1)...,j_d} + h^\tau_{j_1,...,(j_l-1),...,(j_i+1),...,j_d}\right)}{4\Delta_l\Delta_i},\end{aligned} \tag{23}$$

for all space grid points x_j which are not boundary points, i.e. $\forall j_k = 1, ..., N_k$, $\forall k = 1, ..., d$.

At the boundaries of the spacial domain, we introduce the one-sided first order approximations

$$\frac{\partial h}{\partial x_l}(\tau, ..., x_l(0), ...) \approx \frac{h^\tau_{...,j_{l-1},1,j_{l+1},...} - h^\tau_{...,j_{l-1},0,j_{l+1},...}}{\Delta_l}, \tag{24}$$

$$\frac{\partial^2 h}{\partial x_l^2}(\tau, ..., x_l(0), ...) \approx \frac{h^\tau_{...,j_{l-1},2,j_{l+1},...} - 2h^\tau_{...,j_{l-1},1,j_{l+1},...} + h^\tau_{...,j_{l-1},0,j_{l+1},...}}{\Delta_l^2}, \tag{25}$$

$$\frac{\partial h}{\partial x_l}(\tau, ..., x_l(N_l+1), ...) \approx \frac{h^\tau_{...,j_{l-1},N_l+1,j_{l+1},...} - h^\tau_{...,j_{l-1},N_l,j_{l+1},...}}{\Delta_l}, \tag{26}$$

$$\frac{\partial^2 h}{\partial x_l^2}(\tau, ..., x_l(N_l+1), ...) \approx \frac{h^\tau_{...,j_{l-1},N_l+1,j_{l+1},...} - 2h^\tau_{...,j_{l-1},N_l,j_{l+1},...} + h^\tau_{...,j_{l-1},N_l-1,j_{l+1},...}}{\Delta_l^2}. \tag{27}$$

Finally, the initial condition of the pricing function h is given by its payoff (20), which in the FD setting is

$$h_j^0 = H(x_1(j_1), ..., x_d(j_d)) = \left(K - \sum_{i=1}^{n} \omega_i e^{x_{3i-2}(j_{3i-2})}\right)^+,$$

for all $j_k = 0, ..., N_k + 1$ and for all $k = 1, ..., d$.

3.2 Matrix formulation

Now we write the discretization presented above in a matrix format. Dirichlet and von Neumann boundary conditions can be imposed without evaluating the function h at the corresponding boundary point. These conditions can be simply imposed by changing the first and last rows of discretization matrices, as will be explained below. However, this cannot be done for the other boundary conditions we consider (when $x_{3i-1} = 0$ and $x_{3i} = \pm 1$) and we therefore need to include these boundary points in the domain where we solve the option price for. As a consequence, we only solve for the interior points in the y_i dimensions, for the lower boundary and interior points in the v_i dimensions and for interior and both boundary points in the ρ_i dimensions. We can then introduce the number $\tilde{N}_i$, $1 \leq i \leq d$, of points in dimension i for which we compute the solution to (18). We have $\tilde{N}_{3i-2} = N_{3i-2}$, $\tilde{N}_{3i-1} = N_{3i-1} + 1$ and $\tilde{N}_{3i} = N_{3i} + 2$, for all $i \in \{1, ..., n\}$.

We denote by $\underline{h}^\tau$ the vector in $\mathbb{R}^{\prod_{i=1}^d \tilde{N}_i}$, function of τ, such that

$$\underline{h}^\tau \left(\sum_{i=1}^{d} j_i \prod_{k=i+1}^{d} \tilde{N}_k\right) = h^\tau_{j_1,...,j_d},$$

for any $\tau \in [0, T]$ and where indices $j_{3i-2} \in \{1, ..., N_{3i-2}\}$, $j_{3i-1} \in \{0, ..., N_{3i-1}\}$ and $j_{3i} \in \{0, ..., N_{3i} + 1\}$. We use here the convention $\prod_{k=d+1}^{d}(\cdot) := 1$. This indexation of the option price vector h^τ means that the elements of $\underline{h}^\tau$ (from first to last) are ordered such that we are first looping on j_d, then j_{d-1}, etc.

3.2.1 Tensor product representation

Let us define elementary matrices that will enter in the representation of the finite difference operators. Consider an integer $m > 3$. We introduce the $m \times m$ matrices

$$\mathbf{G}_{\mathrm{m}}^{ND} = \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & & & -1 & 0 \end{pmatrix}; \mathbf{G}_{\mathrm{m}}^{\emptyset N} = \begin{pmatrix} -2 & 2 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & & & -1 & 1 \end{pmatrix}; \mathbf{G}_{\mathrm{m}}^{\emptyset\emptyset} = \begin{pmatrix} -2 & 2 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & & & -2 & 2 \end{pmatrix},$$

which correspond to the discretization (up to a scaling factor) of a one-dimensional gradient operator with boundary conditions: D for Dirichlet, N for von Neumann, $\emptyset$ for no explicit boundary conditions but simply that the PDE must be satisfied at the boundary point.

We also introduce the $m \times m$ matrices

$$\mathbf{H}_{\mathrm{m}}^{ND} = \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & & & 1 & -2 \end{pmatrix}; \mathbf{H}_{\mathrm{m}}^{\emptyset N} = \begin{pmatrix} 1 & -2 & 1 & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & & & 1 & -1 \end{pmatrix}; \mathbf{H}_{\mathrm{m}}^{\emptyset\emptyset} = \begin{pmatrix} 1 & -2 & 1 & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & 1 & -2 & 1 \end{pmatrix},$$

which correspond to the discretization (up to a scaling factor) of a one-dimensional Hessian operator with boundary conditions: D for Dirichlet, N for von Neumann, $\emptyset$ for no explicit boundary conditions but simply that the PDE must be satisfied at the boundary point.

We define here the *sorted Kronecker product* to simplify notations in the following.

Definition 3.1. *For some $m \in \mathbb{N}^*$, consider a permutation z on the set of integers $\{1, ..., m\}$,*

$$z : \begin{array}{ccc} \{1, ..., m\} & \longrightarrow & \{1, ..., m\} \\ k & \mapsto & z(k) \end{array}$$

and the matrices $\left(\mathbf{X}_{\tilde{N}_k}\right)_{1 \le k \le m}$. The sorted Kronecker product with increasing indices is then

$$^s\left(\mathbf{X}_{\tilde{N}_{z(1)}} \otimes \cdots \otimes \mathbf{X}_{\tilde{N}_{z(m)}}\right) := \mathbf{X}_{\tilde{N}_1} \otimes \cdots \otimes \mathbf{X}_{\tilde{N}_m}.$$

Using the operator discretizations presented in section 3.1.2, we can write the semi-discretized

operator $\mathcal{A}$ as the matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{L} - r \bigotimes_{j=1}^{d} \mathbf{I}_{\tilde{N}_j}, \tag{28}$$

with

$$\mathbf{L} = \sum_{i=1}^{d} \frac{1}{2\Delta_i} \mathbf{M}_i \mathbf{G}_i + \frac{1}{2} \sum_{i=1}^{d} \mathbf{Q}_{i,i} \frac{1}{\Delta_i^2} \mathbf{H}_{i,i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{k \neq i} \mathbf{Q}_{i,k} \frac{1}{2\Delta_i} \mathbf{G}_i \frac{1}{2\Delta_k} \mathbf{G}_k \tag{29}$$

where $\mathbf{G}_i$ is the tensor corresponding to the dimension-i gradient

$$\mathbf{G}_{3i-2} = {}^{s}\left(\mathbf{G}^{ND}_{\tilde{N}_{3i-2}} \bigotimes_{j \neq 3i-2}^{d} \mathbf{I}_{\tilde{N}_j} \right);\ \mathbf{G}_{3i-1} = {}^{s}\left(\mathbf{G}^{\emptyset N}_{\tilde{N}_{3i-1}} \bigotimes_{j \neq 3i-1}^{d} \mathbf{I}_{\tilde{N}_j} \right);\ \mathbf{G}_{3i} = {}^{s}\left(\mathbf{G}^{\emptyset\emptyset}_{\tilde{N}_{3i}} \bigotimes_{j \neq 3i}^{d} \mathbf{I}_{\tilde{N}_j} \right),$$

$\mathbf{M}_i$ is the tensor corresponding to the discretized drift μ_i defined in (9)-(11)

$$\begin{aligned}
\mathbf{M}_{3i-2} &= {}^{s}\left(\mathrm{Diag}\left\{ r - \frac{1}{2} v_i(j_{3i-1}) \right\}_{j_{3i-1}} \bigotimes_{j \neq 3i-1}^{d} \mathbf{I}_{\tilde{N}_j} \right); \\
\mathbf{M}_{3i-1} &= {}^{s}\left(\mathrm{Diag}\left\{ \kappa_i(\theta_i - v_i(j_{3i-1})) \right\}_{j_{3i-1}} \bigotimes_{j \neq 3i-1}^{d} \mathbf{I}_{\tilde{N}_j} \right); \\
\mathbf{M}_{3i} &= {}^{s}\left(\mathrm{Diag}\left\{ \gamma_i(\rho^i_\infty - \rho_i(j_{3i})) \right\}_{j_{3i}} \bigotimes_{j \neq 3i}^{d} \mathbf{I}_{\tilde{N}_j} \right),
\end{aligned}$$

$\mathbf{H}_{ii}$ is the tensor corresponding to the discretized one-dimensional Hessian in dimension i:

$$\mathbf{H}_{3i-2,3i-2} = {}^{s}\left(\mathbf{H}^{ND}_{\tilde{N}_{3i-2}} \bigotimes_{j \neq 3i-2}^{d} \mathbf{I}_{\tilde{N}_j} \right);\ \mathbf{H}_{3i-1,3i-1} = {}^{s}\left(\mathbf{H}^{\emptyset N}_{\tilde{N}_{3i-1}} \bigotimes_{j \neq 3i-1}^{d} \mathbf{I}_{\tilde{N}_j} \right);\ \mathbf{H}_{3i,3i} = {}^{s}\left(\mathbf{H}^{\emptyset\emptyset}_{\tilde{N}_{3i}} \bigotimes_{j \neq 3i}^{d} \mathbf{I}_{\tilde{N}_j} \right),$$

and $\mathbf{Q}_{i,k}$ corresponds to the discretization of symmetric matrix function Q defined in (12) which we provide in Appendix 2.

Remark 3.1. *We note that by construction $\mathbf{M}_i$ is a rank-1 separable tensor for all $i \in \{1, ..., d\}$. Because of the structure of the correlation function $C_{i,j}$ in (6), $\mathbf{Q}_{i,k}$ can either be rank-1 or rank-2 separable.*

As a consequence, the semi-discretization of the PDE (18) yields the vector ODE

$$\frac{\partial \underline{h}}{\partial \tau}(\tau) - \mathbf{A}\underline{h} = 0, \quad \forall \tau \in (0, T],$$

with the discretized initial condition.

For the time discretization, we introduce $M+1$ times levels $\tau_k = k\Delta\tau$, for $k = 0, ..., M$, where $M \in \mathbb{N}^*$. We denote by $\underline{h}^k$ the approximation of $\underline{h}$ at $\tau = \tau_k$. To solve the ODE, we use a θ time-stepping scheme, i.e., we solve

$$\left(\frac{1}{\Delta\tau}\mathbf{I} - \theta\mathbf{A}\right)\underline{h}^{k+1} = \left(\frac{1}{\Delta\tau}\mathbf{I} + (1-\theta)\mathbf{A}\right)\underline{h}^k \tag{30}$$

for $\theta \in [0,1]$ and $k = 0, ..., M-1$. The parameter θ controls whether the scheme is implicit ($\theta \in (0,1]$) or fully explicit ($\theta = 0$). $\mathbf{I}$ denotes the identity matrix with same size as matrix $\mathbf{A}$.

3.2.2 Tensor Train format

The matrix $\mathbf{A}$ above is of size $\prod_{i=1}^{d} \tilde{N}_i^2$ and the computation of the matrix-vector products above (necessary to find the solution to the vector ODE) has a complexity growing exponentially with d. We therefore need to compress the matrices and vectors or store them in a format that is not subject to the curse of dimensionality or compatible with large values of d.

For this purpose, we introduce the Tensor Train format of Oseledets and Tyrtyshnikov. A n-dimensional tensor $v \in \mathbb{R}^{N_1 \times \cdots \times N_n}$ is said to be in the TT format with *TT ranks* $r_1, ..., r_{n-1}$ if its elements can be represented as

$$v_{i_1,...,i_n} = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{n-1}=1}^{r_{n-1}} U_1(i_1, \alpha_1) U_2(\alpha_1, i_2, \alpha_2) \cdots U_{n-1}(\alpha_{n-2}, i_{n-1}, \alpha_{n-1}) U_n(\alpha_{n-1}, i_n), \tag{31}$$

for all $1 \leq i_k \leq N_k$ and $1 \leq k \leq n$. The arrays U_i are called *TT cores* where $U_1 \in \mathbb{R}^{N_1 \times r_1}$, $U_n \in \mathbb{R}^{r_{n-1} \times N_n}$ and $U_k \in \mathbb{R}^{r_{k-1} \times N_k \times r_k}$ for $2 \leq k \leq n-1$. (31) means that for fixed indices $(i_1, ..., i_n)$ the element $v_{i_1,...,i_n}$ of the tensor v is computed by multiplying a row vector with matrices and the result with a column vector to get a scalar.

The memory storage for the tensor v in the TT-format is $\sum_{i=1}^{n} N_i r_{i-1} r_i \leq r^2 N n$, where we have use the convention $r_0 = r_n = 1$ and notations $r := \max\{r_i\}, N = \max\{N_i\}$. The storage costs therefore scale linearly in the dimension n provided that the TT ranks are bounded. The complexity of basic operations (e.g. matrix-vector product) in this format are usually $O(r^2 N n)$ or $O(r^3 N n)$ which makes it possible to solve PDEs in high dimensions. The TT ranks depend on the ranks of some unfolding matrices of v (see Oseledets (2011), Kolda and Bader (2009)) and approximate low-rank TT tensors can be computed via standard matrix algorithms such as SVD (Singular Value Decomposition) and QR decompositions, making the TT representation robust. To cast matrices $M \in \mathbb{R}^{N_1^2 \times \cdots \times N_n^2}$ into the TT representation (31), the k-th row and column index $(i_k, j_k) \in \{1, ..., N_k\}^2$ is viewed as one long index in $\{1, ..., N_k^2\}$:

$$M(i_1, j_1, ..., i_n, j_n) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{n-1}=1}^{r_{n-1}} U_1(i_1, j_1\alpha_1) U_2(\alpha_1, i_2, j_2, \alpha_2) \times \cdots$$
$$\cdots \times U_{n-1}(\alpha_{n-2}, i_{n-1}, j_{n-1}, \alpha_{n-1}) U_n(\alpha_{n-1}, i_n, j_n),$$

To increase even more the efficiency of the TT representation for a vector/matrix, Khoromskij (2011) and Oseledets (2009) have introduced further tensorizations leading to the *Quantized* tensor train representation. For instance, assume that the vector v in (31) has indices $i_k \in \{1, ..., 2^{l_k}\}$, i.e., $N_k = 2^{l_k}$. We can represent the scalar index i_k by a multi-index $(i_{k,1}, ..., i_{k,l_k})$ via its binary encoding $i_k = 1 + \sum_{j=1}^{l_k} 2^{l_k - j}(i_{k,j} - 1)$ where each $i_{k,j}$ takes values 1 or 2. The k-th dimension of the vector v is now represented by l_k virtual dimensions. This means that v is now viewed as a $l_1 + ... + l_n$ dimensional $\underbrace{2 \times ... \times 2}_{l_1} \times ... \times \underbrace{2 \times ... \times 2}_{l_n}$ tensor. Quantization is therefore a reshaping algorithm for tensors that increases the number of dimensions. Other quantizations are possible, however, this is the quantization generally used in order to extract as much QTT structure as possible and compress the data as much as possible.

A *low rank* for the (Q)TT representation of tensors is the key to efficient compression of huge vectors/matrices. Exact low rank (Q)TT representations are known for many matrices and vectors used in the discretization of high dimensional PDEs, see Kazeev and Khoromskij (2012) for the Laplace operator and more generally Kazeev, Reichmann, and Schwab (2013) for linear diffusion operators. In cases where an exact low rank representation is unknown, it is possible to approximate a given tensor in the (Q)TT format for a chosen level of accuracy (in the Frobenius norm).

In the following, we will compute exact (whenever possible) or approximate QTT representations of the vectors $\underline{h}^k$ and matrix $\mathbf{A}$ entering in the finite difference discretization (30). We use the TT Matlab toolbox publicly available at `http://spring.inm.ras.ru/osel/?page_id=24` and in particular the TT-solver routine *dmrg_solve2* to approximate the solution to the linear system (30).

4 Application: Pricing index options

Providing protection against financial crashes and systemic risk, equity index put options are one of the most traded derivatives on financial exchanges. In this section, we investigate the impact of stochastic correlation on the price of index put options. In particular, we are interested in the ability of stochastic correlation to generate steep implied volatility smiles for the index, or equivalently to increase the probability of large negative returns in the index returns' distribution.

We first briefly discuss why the number of common factors should be at least two when modelling three or more assets. Then, we detail the index option pricing application.

4.1 Number of common factors

The remark below explains why it is important to have at least two common factors when modelling three or more assets.

Remark 4.1. *Only one common factor.*
In this framework, it is imposed that there is at least one common factor $\bar{W}^0$ driving all asset correlations and one idiosyncratic factor $\hat{W}^i$. When $\bar{W}^0$ is the only common factor, i.e. when $\alpha_k^i = 0$ for $k = 1, ..., p$ (hence $\beta^i = 1$), we have the following correlation structure

$$C_{ii} = 1 \quad and \quad C_t^{ij} = \rho_t^i \rho_t^j \ , \quad i \neq j. \tag{32}$$

Even though this correlation structure is valid, it is easy to see that it is very restrictive. Assume there are three assets. Then, the model imposes the following structure:

$$Correl_t\left(dY_t^1, dY_t^2\right) > 0 \quad and \quad Correl_t\left(dY_t^1, dY_t^3\right) < 0 \quad \Rightarrow \quad Correl_t\left(dY_t^2, dY_t^3\right) < 0.$$

The way to avoid this restrictive structure is to have at least a second common factor $\bar{W}^1$. As a consequence, the model will not impose any particular sign for $Correl_t\left(dY_t^2, dY_t^3\right)$, since C_t^{ij} then becomes $C_t^{ij} = \rho_t^i \rho_t^j + \sqrt{1 - \left(\rho_t^i\right)^2}\sqrt{1 - \left(\rho_t^j\right)^2}\alpha_1^i \alpha_1^j$, $i \neq j$.

In all generality, one could take n common factors $\{\bar{W}_k\}_{k=1,...,n}$ and still have $O(n)$ stochastic factors but the number of parameters will be $O(n^2)$. Even though this makes the model more flexible, there is a always a trade-off between increasing the complexity of the model and having a fast estimation procedure with parameters easily identifiable. As a consequence, we recommend to have a number of correlation processes bounded and not increasing with n.

4.2 Pricing Equity index options

To illustrate our framework, we consider a multivariate Heston model where the correlation between stocks is stochastic. The dynamics are given by (1) - (2) - (3) - (5) with

$$b_i(x) = \nu_i \sqrt{x}, \quad (\nu_i^2 < 2\kappa_i \theta_i). \tag{33}$$

Given that the number of assets is not relevant to show the impact of stochastic correlation, we simply consider $n = 2$ assets which corresponds to solving a $6 + 1$-dimensional PDE. We consider an index with weights $\omega_1 = 0.6, \omega_2 = 0.4$ and $S_0^1 = 40, S_0^2 = 60$, the spot index value therefore being 48. We now construct the smile of volatility from index put options with maturities $T = 2/12, 3/12, 6/12$ years and strikes between $K = 15$ and $K = 65$ (with smaller ranges of strikes for shorter maturities as in the financial markets).

For the individual assets' parameters, we consider values in the range of what can be obtained from the calibration of the Heston model to options (with the Feller condition being satisfied). We choose for this example:

$$\kappa_i = 4.2; \theta_i = 0.08; \nu_i = 0.60; \zeta_i = -0.85.$$

Other choices of parameters lead to the same qualitative conclusions.

For the correlation processes, we choose a parameter set which corresponds to the case where the asset returns and the correlation between assets are strongly negatively correlated.

$$\gamma_i = 2; \rho_\infty^i = 0.6; \xi_i = 0.85; \delta_i = -0.9.$$

For the discretization of each spacial variable, we have the following number of degrees of freedom for each asset:

$$\tilde{N}_y = 2^{10}; \tilde{N}_v = 2^9; \tilde{N}_\rho = 2^8,$$

and for the time stepping, we use a Crank-Nicholson scheme, i.e. $\theta = 1/2$, with $M = 30$ time steps.

When increasing the number of assets, it turns out that the QTT ranks of the matrix $\mathbf{A}$ stay more or less constant (around 15 on average, even for 10 assets). However, the ranks of the approximation to the option price $\underline{h}^T$ grow with n which renders the computations unfeasible at some point (between 15 and 20 dimensions depending on the model specifications) and one therefore needs to resort to Monte Carlo methods to find the option price. In fact, the ranks heavily depend on the accuracy we require from the TT solver when solving the system (30). For instance, for two assets (6 dimensions) and a tolerance of $tol =$ 1e-8, the QTT ranks of the solution are around 150 and it can take up to 15 hours to compute the option price on a dual core 2.2GHz with 4Go of RAM. However, when $tol =$ 5.0e-6, ranks are as low as 20 and the computation takes 90 seconds (and for 4 assets (12 dimensions), the ranks are around 55 and the computational time is 1.5 hours). Monte-Carlo simulations are much faster if we want to evaluate the option price only at one particular point, but if one needs to estimate the model and evaluate the model for many different initial values of volatility and correlations, the QTT - PDE pricer is significantly much faster.

In Figure 16, we have plotted the implied volatility smiles generated by a constant correlation model denoted by SVCC (stochastic volatility constant correlation) and by a stochastic correlation model denoted by SVSC (stochastic volatility stochastic correlation). Fixing one particular maturity, we see that adding stochastic correlation to the model increases the steepness of the implied volatilities, all other parameters remaining the same. For instance, the 2 months maturity slices intersect around $K = 55$ while the SVSC model generates 5 to 10% higher implied volatilities for smaller strikes $K = 30$. This feature seems to appear regardless of the initial value of variance and correlation we considered.

Given that the shape of implied volatilities are directly related to the distribution of returns under a

risk neutral measure, we expect the distribution of the SVSC model for negative returns to be heavier tailed than that of the SVCC model. In Figure 17, one can see the impact of stochastic correlation on the left tail of the returns' distribution for a one year horizon. In particular, when we estimated the probability of a 70% drop in the returns over one year, the probability with stochastic correlation was around 10 times higher than that of the constant correlation model (for the set of parameters given above).

As a consequence, we find that stochastic correlation can generate steeper implied volatility smiles for indices and higher probabilities of large negative returns than constant correlation models.

Appendix

1 Existence and uniqueness of a strong solution

In this section, we state theorem 1.1 of Luo (2011) which we use to prove existence of a unique strong solution to the SDEs presented in this article.

Consider the SDE

$$dX_t = \eta(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x, \tag{34}$$

where $\eta : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathcal{M}_{d\times m}$ are respectively a vector and a matrix valued function. B is a standard Brownian motion on $\mathbb{R}^m$.

In the following, $||\ ||_2$ will denote the Euclidean norm of a vector and $||\ ||_F^2$ the Frobenius norm of a matrix.

Assumption 1.1. *There exists a strictly increasing function $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $f(0)$ and $\int_{0^+} f^{-2}(u)du = \infty$, such that for all $i \in \{1, ..., d\}$ and all $x, y \in \mathbb{R}$,*

$$||\sigma(x)^{i,\cdot} - \sigma(y)^{i,\cdot}||_2 \leq f\left(\sqrt{\sum_{k=1}^{i} |x_k - y_k|^2}\right)$$

where $\sigma(x)^{i,\cdot}$ denotes the i-th row of the matrix $\sigma(x)$.

Assumption 1.2. *There exists a strictly increasing concave function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $g(0)$ and $\int_{0^+} g^{-1}(u)du = \infty$, such that for all $i \in \{1, ..., d\}$ and all $x, y \in \mathbb{R}$,*

$$|\eta_i(x) - \eta_i(y)| \leq g\left(\sqrt{\sum_{k=1}^{i} |x_k - y_k|^2}\right).$$

where η_i is the i-th component of the vector η.

Assumption 1.3. *There exists a constant $c > 0$ such that for all $x \in \mathbb{R}$,*

$$||\eta(x)||_2^2 + ||\sigma(x)||_F^2 \leq c\left(1 + ||x||_2^2\right).$$

Theorem 1.1. Existence of a solution to the SDE (34)**.**
Under the assumptions 1.1, 1.2 and 1.3, there exists a unique strong solution to the SDE (34).

The proof of theorem 1.1 can be found in Luo (2011).

Remark 1.1. *The assumption 1.1 can be localized and changed to the following. Consider the notation* $B(R)$ *for the closed ball* $B(R) := \{x \in \mathbb{R}^d, ||x||_2 \leq R\}$ *for some* $R > 0$.

There exists a constant $c_0 > 0$ *and a strictly increasing function* $f : (0, c_0] \mapsto \mathbb{R}_+$ *with* $f(0)$ *and* $\int_0^{c_0} f^{-2}(u)du = \infty$, *such that*

$$\forall R > 0, \exists C_R > 0, \forall i \in \{1, ..., n\}, \forall x, y \in B(R),$$
$$||x - y|| \leq c_0 \Longrightarrow ||\sigma(x)^{i,\cdot} - \sigma(y)^{i,\cdot}||_2 \leq C_R f\left(\sqrt{\sum_{k=1}^{i} |x_k - y_k|^2}\right)$$

2 Tensor representation of the discretized covariance matrix

We detail here the entries of the tensor $\mathbf{Q}_{i,k}$ (entering equation (29)) which corresponds to the discretization of covariance matrix function Q defined in (12)-(13)-(14). Section 3.1 details the discretization step and section 3.2 introduces the notations for the matrix formulation.

The tensor $\mathbf{Q}_{i,k}$ corresponds to the discretization of function $x \mapsto Q_{i,k}(x)$ which can always be written in the form

$$Q_{i,k}(x) = \sum_{r=1}^{\bar{r}} c_{i,k}^{r} \times q_{i,k}^{r,1}(x_1) \times \cdots \times q_{i,k}^{r,d}(x_d), \tag{35}$$

where $\bar{r} \leq 2$. We stress that each $q_{i,k}^{r,j}$ is a *one-dimension function* (for all $j = 1, ..., d$) and c_r is a constant for each $(i, k, r) \in \{1, ..., d\}^2 \times \{1, ..., \bar{r}\}$.

For example, $Q_{3i,3i-2}(x) = \sqrt{x_{3i-1}} x_{3i} \xi_i \sqrt{1 - x_{3i}^2} \delta_i$ can be written as above with the specifications

$$\begin{aligned}
\bar{r} &= 1, \\
c_{3i,3i-2} &= \xi_i \delta_i, \\
q_{3i,3i-2}^{3i-1}(x_{3i-1}) &= \sqrt{x_{3i-1}}, \\
q_{3i,3i-2}^{3i}(x_{3i}) &= x_{3i}\sqrt{1 - x_{3i}^2} \\
q_{3i,3i-2}^{j}(x_j) &= 1 \ , \quad \text{otherwise.}
\end{aligned}$$

In fact, all functions $Q_{i,k}(x)$ that do not depend on the correlation function C_{ik} (or C_{ki}) satisfy $\bar{r} = 1$ whereas functions $Q_{i,k}(x)$ that depend on C_{ik} (or C_{ki}) satisfy $\bar{r} = 2$. For example, $Q_{3i-2,3j-2}(x)$ can be written as above with the specifications

$$\begin{aligned}
\bar{r} &= 2, \\
c_{3i-2,3j-2}^{1} &= 1, & c_{3i-2,3j-2}^{2} &= \left(\alpha^i\right)^\top \alpha^j, \\
q_{3i-2,3j-2}^{1,3i-1}(x_{3i-1}) &= \sqrt{x_{3i-1}} \ , & q_{3i-2,3j-2}^{1,3j-1}(x_{3j-1}) &= \sqrt{x_{3j-1}}, \\
q_{3i-2,3j-2}^{1,3i}(x_{3i}) &= x_{3i}, & q_{3i-2,3j-2}^{1,3j}(x_{3j}) &= x_{3j}, \\
q_{3i-2,3j-2}^{2,3i-1}(x_{3i-1}) &= \sqrt{x_{3i-1}}, & q_{3i-2,3j-2}^{2,3j-1}(x_{3j-1}) &= \sqrt{x_{3j-1}}, \\
q_{3i-2,3j-2}^{2,3i}(x_{3i}) &= \sqrt{1 - x_{3i}^2}, & q_{3i-2,3j-2}^{2,3j}(x_{3j}) &= \sqrt{1 - x_{3j}^2}.
\end{aligned}$$

Since all coefficients $Q_{i,k}$ satisfy the decomposition (35), the tensor representation of the discretized

function have the same form and are at most rank-$\bar{r}$ separable:

$$\mathbf{Q}_{i,k} = \sum_{r=1}^{\bar{r}} c_{i,k}^{r} \times \mathbf{q}_{i,k}^{r,1} \otimes \cdots \otimes \mathbf{q}_{i,k}^{r,d},$$

where $\mathbf{q}_{i,k}^{r,l}$ is a diagonal matrix (for all $l = 1, ..., d$) where the diagonal elements are the function $q_{i,k}^{r,l}$ evaluated at the grid points $\{x_l(j_l)\}_{j_l}$ in the l-th dimension.

3 Figures

Figure 16: Implied volatility smiles generated by a stochastic volatility constant correlation model (SVCC) and a stochastic volatility stochastic correlation model (SVSC). The parameters of the models are given in section 4.2 and the smiles plotted correspond to maturities of 2,3 and 6 months.

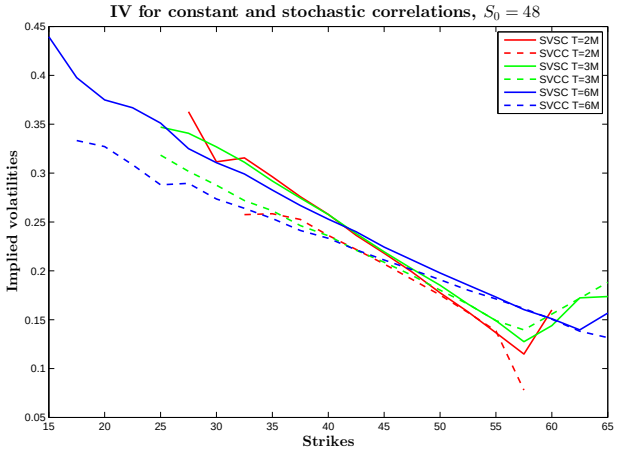
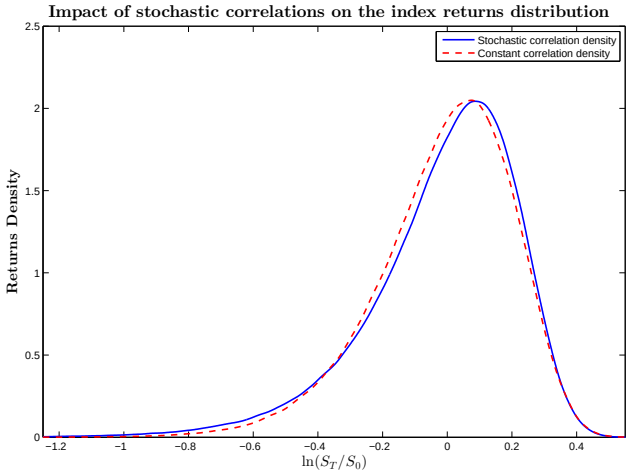


Figure 17: Risk neutral distribution of returns over a one year horizon in the case of constant (SVCC model) and stochastic correlation (SVSC). The parameters for these models are reported in section 4.2.



REFERENCES

Ahdida, A., and A. Alfonsi, 2013, "A mean-reverting SDE on correlation matrices," *Stochastic Processes and their Applications*, 123(4), 1472–1520.

Alfonsi, A., 2010, "High order discretization schemes for the CIR process: application to affine term structure and heston models.," *Math. Comput.*, 79(269), 209–237.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge, 1988, "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96(1), pp. 116–131.

Bru, M., 1991, "Wishart Processes," *Journal of Theoretical Probability*, 4, 725–751.

Cont, R., and R. Deguest, 2010, "Equity Correlations Implied by Index Options: Estimation and Model Uncertainty Analysis," *Mathematical Finance*.

Cuchiero, C., D. Filipovic, E. Mayerhofer, and J. Teichmann, 2009, "Affine Processes on Positive Semidefinite Matrices," working paper, ETH Zurich, manuscript.

Da Fonseca, J., M. Grasselli, and C. Tebaldi, 2007, "Option pricing when correlations are stochastic: an analytical framework," *Review of Derivatives Research*, 10(2), 151–180.

Driessen, J., P. J. Maenhout, and G. Vilkov, 2009, "The Price of Correlation Risk: Evidence from Equity Options," *Journal of Finance*, 64(3), 1377–1406.

Duffy, D. J., 2006, *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach (The Wiley Finance Series)*. Wiley.

Ethier, S. N., and T. G. Kurtz, 1986, *Markov Processes: Characterization and Convergence (Wiley Series in Probability and Statistics)*. Wiley.

Gourieroux, C., and J. Jasiak, 2006, "Multivariate Jacobi process with application to smooth transitions," *Journal of Econometrics*, 131.

Gourieroux, C., J. Jasiak, and R. Sufana, 2009, "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150(2), 167 – 181.

Guyon, J., 2013, "A New Class of Local Correlation Models," *SSRN eLibrary*.

Kaya Boortz, C., 2008, "Modelling correlation risk.," *Diplomarbeit, preprint*.

Kazeev, V., and B. Khoromskij, 2012, "Low-Rank Explicit QTT Representation of the Laplace Operator and Its Inverse," *SIAM Journal on Matrix Analysis and Applications*, 33(3), 742–758.

Kazeev, V., O. Reichmann, and C. Schwab, 2013, "Low-rank tensor structure of linear diffusion operators in the {TT} and {QTT} formats," *Linear Algebra and its Applications*, 438(11), 4204 – 4221.

Khoromskij, B., 2011, "O(dlog N)-Quantics Approximation of N-d Tensors in High-Dimensional Numerical Modeling," *Constructive Approximation*, 34(2), 257–280.

Kolda, T. G., and B. W. Bader, 2009, "Tensor Decompositions and Applications," *SIAM Rev.*, 51(3), 455–500.

Langnau, A., 2010, "A dynamic model for correlation," *Risk magazine*.

Leippold, M., and F. Trojani, 2008, "Asset Pricing with Matrix Jump Diffusions," *SSRN eLibrary*.

Luo, D., 2011, "Pathwise uniqueness of multi-dimensional stochastic differential equations with Hölder diffusion coefficients," *Frontiers of Mathematics in China*, 6(1), 129–136.

Ma, J., 2009, "Pricing Foreign Equity Options with Stochastic Correlation and Volatility," *Annals of Economics and Finance*.

Moskowitz, T. J., 2003, "An Analysis of Covariance Risk and Pricing Anomalies," *Review of Financial Studies*, 16(2), 417–457.

Oseledets, I., 2009, "Approximation of matrices with logarithmic number of parameters," *Doklady Mathematics*, 80(2), 653–654.

Oseledets, I. V., 2011, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, 33(5), 2295–2317.

Oseledets, I. V., and E. E. Tyrtyshnikov, 2009, "Recursive decomposition of multidimensional tensors," *Doklady Math.*, 427(1), 14–16.

van Emmerich, C., 2006, "Modelling Correlation as a Stochastic Process," working paper, Working paper.

van Haastrecht, A., and A. A. J. Pelsser, 2010, "Efficient almost exact simulation of the Heston stochastic volatility model," *International Journal of Theoretical and Applied Finance (IJTAF)*, 13(01), 1–43.

Veraart, A., and L. Veraart, 2012, "Stochastic volatility and stochastic leverage," *Annals of Finance*, 8(2-3), 205–233.

Model calibration to marginal distributions

Chris Bardgett

Abstract

In this paper, we present an alternative choice of loss function for the calibration of option pricing models. Given the cross section of options' data, we build market implied risk-neutral distributions (RNDs) to which we calibrate the model by minimizing a distance criterion between market and model RNDs. We study the advantages and drawbacks of adding distributional assumptions to the options' data for in and out of sample pricing performance, depending on how many strikes are traded. We find that when the number of traded strikes is small, the calibration to the market distributions can lead to smaller out of sample pricing errors but not systematically. On the other hand, when many strikes are traded, even though RND calibration produces in and out of sample prices within the bid-ask spread, calibration to option prices yields smaller errors. We also show that the RND calibration has the same complexity as the calibration to option prices.

KEYWORDS: model calibration, loss function, market implied risk-neutral distributions, option pricing.

1 Introduction

Since the framework of Black and Scholes (1973) was introduced, many efforts have been made in finance to improve models in order to better account for the stylized facts of asset returns. These include local volatility models as in Dupire (1994), stochastic volatility models (e.g. Heston (1993)), models driven by Lévy processes (e.g. Madan, Carr, and Chang (1998)), and multi-factor models (e.g. Gourieroux, Jasiak, and Sufana (2009)) among many others. A preliminary but very challenging step, before using these models, is to find parameter estimates that fit market data. Calibration to option prices by minimizing absolute pricing errors and relative pricing errors or calibration to implied volatilities has been extensively used (Bakshi, Cao, and Chen (1997), Heston and Nandi (2000) among others). In fact, Christoffersen and Jacobs (2004) have studied the impact of the choice of loss function on the performance of option pricing models and find that using the same loss function when estimating and evaluating the pricing performance of the model is essential. They use options data on the S&P 500 index in order to test the in and out of sample performance of a model. It is clear that regardless of whether a model is misspecified or not, minimizing the option pricing errors - when estimating it - will lead to the smallest in-sample option pricing errors possible across loss functions. However, it is not clear that for out of sample performance, it is necessary to use the same loss function for estimation and performance evaluation. In this paper, we study the impact of choosing different loss functions on the out of sample performance of pricing models by introducing a new loss function based on the minimization between market implied risk-neutral distributions (RNDs) and model RNDs. In order to avoid having results that are influenced by model misspecification, we choose to work with simulated option prices from a reference model and work with these option prices as if they were market data.

For a given maturity, if we assume that vanilla prices are traded for a continuum of strikes, the information contained in those prices is the same as that of the risk neutral marginal distribution (we refer to marginal distributions as the conditional distributions given today's information) of the underlying stochastic process for that maturity (Breeden and Litzenberger (1978)). It therefore makes sense to see calibration to market vanilla prices as an implicit calibration of the underlying stochastic process marginals to the market implied marginals. In practice, there are most of the time only a few strikes traded for a given maturity, but if one imposes additional information on the structure on the option prices traded, it is possible to recover market RNDs. In this paper we will study whether the additional distributional assumptions one can make when constructing an RND from only a few option prices enables to have a better estimation of the stochastic process. In particular we will analyze whether the out of sample performance of the option pricing model improves, i.e., whether using RNDs makes it possible to better identifiability of option pricing parameters.

Constructing the RNDs is a very important stream of research as its results give insights into which distributional assumptions option pricing models should have to reproduce market option prices. It is possible to approximate RNDs using parametric functions or non-parametric statistics. The non-

parametric estimates[45] enable to make very few assumptions to recover the RND, but they usually require a lot of data and are therefore not appropriate to approximate the RND for a slice of options' data with only a few strikes, say between five and ten. On the other hand parametric techniques[46] make assumptions on the structure of the data but can approximate the market RND with only a few strikes traded. We therefore choose to use parametric methods to recover the RNDs for option prices and in particular we choose to use the mixture of log-normal approach introduced by Alexander and Narayanan (2001), Brigo and Mercurio (2002) and very well detailed in Rebonato (2004). This method is relatively easy to implement, very fast and we give examples to show that it is extremely accurate at reproducing option prices traded on the market.

Once the RNDs are estimated, one can choose to construct a stochastic process that has the specified marginal distributions[47]. In practice, one may not desire to fit exactly the market RNDs as there may not be enough information on the market to estimate the whole RND (especially far in the tails) precisely. Another reason is that for option pricing and hedging applications, one may want to have closed form option prices or greeks that can be calculated easily. For this reason, in this paper, we postulate a model and then calibrate the model to the market RNDs by minimizing some distance between the model and market RNDs. This way, we can choose to put more emphasis on fitting one specific part of the market RND (for example if there is a high uncertainty for one tail of the market RND, we will can put less emphasis on its fitting). Additionally, if the market RND is attainable, i.e. there exists a set of model parameters for which the model is exactly the market RND, then the model will have the exact marginals. Different criteria have been proposed in the literature when it comes to minimizing the distance between distributions, in particular using the relative entropy (Avellaneda (1998), Cont and Tankov (2004b), Cont and Tankov (2006)). Based on the relation between option prices and cdfs as well as the uncertainty of the market RND in the tails, we propose to minimize a weighted L^p distance between survival cdfs.

Option pricing models do not usually have closed forms for the distribution of the underlying[48]. However, it is possible to approximate the model RNDs very efficiently using the same techniques one would use to approximate option prices. For instance, models that have a tractable characteristic function for the returns have densities and cumulative distribution functions (cdfs) in semi closed-form via Fourier inversion. These include most of the models used in practice, for instance the affine class of Duffie, Pan, and Singleton (2000) or Lévy processes (see Cont and Tankov (2004a)). "Non-tractable" models that do not have a closed-form distribution or characteristic function for the returns can be simulated by Monte-Carlo methods and the distribution of returns can be easily approximated by

[45] We refer to Jackwerth and Rubinstein (1996), Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2003) and Fan and Mancini (2009) for techniques to recover non-parametrically the market RNDs.

[46] We refer to Bahra (1997), Bliss and Panigirtzoglou (2004), Bu and Hadri (2007) and the book Rebonato (2004) to name only a few.

[47] This is the approach of Madan and Yor (2002) who construct Markovian processes with one specific target distribution, in particular the local volatility model of Dupire (1994). Other references include Gyöngy (1986), Bentata and Cont (2009).

[48] Some exceptions include for instance Lévy process such as the variance gamma process of Madan, Carr, and Chang (1998), etc. In these cases, it is much faster to calibrate the model via RND calibration than via calibration to option prices.

smoothing the empirical cdf. Overall the complexity of computing model RNDs is the same as that of computing option prices. To our knowledge, estimating a model to a market distribution was first performed indirectly by Andersen and Brotherton-Ratcliffe (1998) to obtain stable calibrated local volatilities using PDE methods, and later in a different context (in an equilibrium setting) by Ziegler (2007). Here we develop this technique for option pricing applications and analyze in detail its advantages and drawbacks.

To test the RND calibration, we simulate option prices from a known model - in particular we present results for the Heston model. Then we estimate this model - assuming we do not know its parameters - in two different ways: minimizing the distance between market and model option prices or minimizing the distance between market and model cdfs. Then with both sets of parameters, we test the in sample and out of sample pricing performance of the model on option prices the model was not estimated to. We find that the RND calibration performs very well in sample (i.e. resulting option prices are well within the bid-ask spread) whether there are only a few strikes (say, 5) or many, as long as we do not impose the model to fit the market RND where there is absolutely no data, i.e. far in the tails. The mixture of log-normal distributions is very precise at recovering the market RND even when only a few data points are available, but the extrapolation far in the tails (where there is no data) eventually fails due to the log-normal assumption. Out of sample, the evidence is mixed and option prices errors (or implied volatility errors) can be better for the RND calibration or option prices calibration depending on the market situation, i.e. on the parameters of the model.

The remaining of the paper is organized as follows. Section 2 introduces the RND calibration method and suggests different ways to calibrate an option pricing model to given market marginal distributions for the underlying process. We propose different criteria to match model and market distributions and relate RND calibration to calibration to option prices. Section 3 presents an efficient technique to recover the market RND and proposes different ways to approximate the model RND. Finally, section 4 presents our results where we compare the in and out of sample pricing performance of the RND and option price calibrations.

2 The RND calibration method

In this section, we introduce the RND calibration technique. We assume that we are at time $t = 0$ and that arbitrage free dynamics (under some risk-neutral measure) have been specified for the stock price process $\{S_t\}_{t\geq 0}$ generating the option prices on the market. We further assume that we have an approximation $\hat{f}_t$ (respectively $\hat{F}_t$) of the marginal density (resp. marginal cdf) of the random variable S_t at time $t \leq T$ [49] for some fixed $T > 0$ (see section 3). We denote by g_t the model density (resp. G_t the model cdf) of S_t, for $0 \leq t \leq T$, and by Θ the set of parameters these distributions depend

[49] In practice, we only have one or a couple of the marginal distributions $\hat{f}_{T_1}, ..., \hat{f}_{T_M}$ for some future dates $T_1, ..., T_M$.

on. Unless we specify very particular dynamics for the process S (see for instance Madan and Yor (2002)), there is no reason why the distribution g_t should be flexible enough to exactly replicate $\hat{f}_t$.[50] As a consequence, we need to define distance criteria between $\hat{f}$ and g (resp. between $\hat{F}_t$ and G) to "calibrate" the model. The first criterion we suggest is based on distances in the L^p space and the second one on the relative entropy between densities. Finally, we explain how to relate calibration to option prices and calibration to survival cdfs.

2.1 RND calibration via the L^p norm

We first introduce a distance criterion between the probability density functions $\hat{f}$ and g assuming that they belong to $L^p(\mathbb{R}_+, \mathbb{R}_+)$ for some $p \geq 1$. Given that $\hat{f}$ and g are density functions, they are integrable and at least belong to $L^1(\mathbb{R}_+, \mathbb{R}_+)$.

Let us define a weighting probability measure w on $[0,T] \times \mathbb{R}_+$ giving more or less importance to the market densities $\hat{f}_t$ for $t \in [0,T]$. We define the distance d^p_ω between $\hat{f}$ and g by:

$$d^p_\omega(\hat{f}, g) = \left(\int_{t\in[0,T]} \int_{s\in\mathbb{R}_+} \left| \hat{f}_t(s) - g_t(s) \right|^p \omega(ds, dt) \right)^{1/p}. \tag{1}$$

Similarly, assuming that the survival cdfs $\hat{\bar{F}}$ and $\bar{G}$ are in $L^p(\mathbb{R}_+, [0,1])$ for some $p \geq 1$, we can define

$$d^p_\omega(\hat{\bar{F}}, \bar{G}) = \left(\int_{t\in[0,T]} \int_{s\in\mathbb{R}_+} \left| \hat{F}_t(s) - G_t(s) \right|^p \omega(ds, dt) \right)^{1/p}. \tag{2}$$

In all models used in practice, returns and stock prices have finite expectation (otherwise option prices are not guaranteed to be finite) and therefore the survival cdf at least belongs to L^1.

In practice, the weight function ω is chosen to be $\sum_i^M w_i(ds)\delta_{T_i}(dt)$ because we will only have a finite number of market marginal distributions. The weights w_i are density functions on $\mathbb{R}_+$.

When calibrating option pricing models, we are interested in minimizing one of these distance criterion as in Christoffersen and Jacobs (2004). This inverse problem can be ill-posed and one may resort to regularization techniques to solve the optimization problem. We refer to Cont and Tankov (2004b, 2006, 2004a) for more details.

[50] For instance, assume we specify a Black-Scholes model for the process S and the densities $\hat{f}_t$ are not log-normal, it will not be possible to find parameters to match these densities exactly.

2.2 RND calibration via relative entropy

A popular pseudo-distance criterion for densities is the Kullback-Leibler divergence. For the densities $\hat{f}$ and g, as long as $\hat{f}_t(s) = 0 \Rightarrow g_t(s) = 0$, we can define the weighted relative entropy by

$$d^{re}_{\omega}(\hat{f}, g) = \int\limits_{t\in[0,T]} \int\limits_{s\in\mathbb{R}_+} g_t(s) \ln\left(\frac{g_t(s)}{\hat{f}_t(s)}\right) ds\ \omega(dt). \tag{3}$$

In practice, the weight is simply a sum $\sum_i^M \delta_{T_i}(dt)$ so that $d^{re}_{\omega}(\hat{f}, g) = \sum_i^N d^{re}_{\omega}(\hat{f}_{T_i}, g_{T_i})$

The relative entropy between two densities is non-negative and equal to zero only if the densities are identical. The logarithm emphasizes the fitting of the tails of the distributions compared to the L^p distances presented in section 2.1. The multiplication by g of the logarithm provides control so that the difference $\ln(g) - \ln(f)$ doesn't explode (since $\lim_{x\to 0} x\ln(x) = 0$).

2.3 Link between calibration to options prices and RND calibration

It is possible to see the usual calibration procedure which consists in fitting the market option prices as a particular way to fit a model distributions to the market implied marginals. Let us assume that there are several maturities T_j, $j \in \{1, ..., M\}$ for which there are call options quoted with strikes K_i for $i \in \{1, ..., N_j\}$ depending on maturity T_j.

We use the same notation as before. Additionally, $\hat{f}_{T_j}$ represents the estimator of the risk neutral density of the stock price for maturity T_j and $\hat{F}_{T_j}$ the corresponding cdf. By construction, we have

$$\int\limits_{\mathbb{R}_+} (s - K_i)^+ \hat{f}_{T_j}(s) ds \approx C_i \qquad \forall i \in \{1, ..., N_j\}$$

where $(s - K_i)^+ := \max(s - K_i, 0)$.

We can therefore write the problem of calibrating to call prices (via least squares) as finding the optimal value θ^* of model parameters such that:

$$\begin{aligned}\theta^* &= \arg\min_{\theta}\{\sum_{j=1}^{M}\sum_{i=1}^{N_j}\left|\int\limits_{\mathbb{R}_+} g_{T_j}(s;\theta)(s-K_i)^+ ds - \int\limits_{\mathbb{R}_+} \hat{f}_{T_j}(s)(s-K_i)^+ ds\right|^2\} \\ &= \arg\min_{\theta}\{\sum_{j=1}^{M}\sum_{i=1}^{N_j}\left|\int\limits_{\mathbb{R}_+} (g_{T_j}(s;\theta) - \hat{f}_{T_j}(s))(s-K_i)^+ ds\right|^2\}\end{aligned}$$

So the resulting densities of the traditional calibration procedure are densities $g_{T_j}(\theta^*)$ that match

the scalar product of $\hat{f}_{T_j}$ and the family of functions $s \to \max(s - K_i, 0)$ for all $i \in \{1, ..., N_j\}$. For each maturity T_j, as $N_j \to \infty$ and all strikes $K_i \geq 0$ are considered, the family of functions $\{s \to \max(s - K_i, 0)\}_{i \geq 1}$ becomes dense in the space of continuous function from $\mathbb{R}_+$ to $\mathbb{R}$ and the estimated $g_{T_j}(\theta^*)$ converges at almost every point to $\hat{f}_{T_j}$ (provided that $\hat{f}_{T_j}$ is in the span of the parametric densities $\{g_{T_j}(\theta)\}_{\theta \in \Theta}$). As the number of strikes increase for each maturity, calibration to option prices or to RNDs (minimizing (1) or (3)) yield the same parameter. The differences in calibrated parameters appear when the sample is finite.

It is also possible to consider the calibration problem in terms of the cdfs. Indeed, for most cdf in practice, we can integrate by parts the call price integral

$$\int_{\mathbb{R}_+} (s - K_i)^+ \hat{f}_{T_j}(s) ds = \underbrace{[(s - K_i)^+ (-\hat{\bar{F}}_{T_j}(s))]_0^\infty}_{=0} + \int_{\mathbb{R}_+} 1_{[K_i, \infty)} \hat{\bar{F}}_{T_j}(s) ds$$

where $\hat{\bar{F}}_{T_j}(s) := 1 - \hat{F}_{T_j}(s)$ is the estimated survival cdf. The term in between brackets is zero provided the tails of the distribution are not too fat [51]. The minimization problem can then be written as

$$\begin{aligned} \theta^* &= \arg\min_\theta \{ \sum_{j=1}^{M} \sum_{i=1}^{N_j} \left| \int_{K_i}^{\infty} \bar{G}_{T_j}(s; \theta) ds - \int_{K_i}^{\infty} \hat{\bar{F}}_{T_j}(s) ds \right|^2 \} \\ &= \arg\min_\theta \{ \sum_{j=1}^{M} \sum_{i=1}^{N_j} \left| \int_{K_i}^{\infty} [\bar{G}_{T_j}(s; \theta) - \hat{\bar{F}}_{T_j}(s)] ds \right|^2 \} \end{aligned} \tag{4}$$

If there are only a finite number of strikes, this criterion will clearly not fit the distributions as well as (2). In fact, it can be the case that the minimum (4) is zero whereas the distance (2) is non-zero and not minimal, i.e. it is possible that all available option prices are replicated but the distributions still differ. Indeed, the price of a call option is merely the integral over a subset of $\mathbb{R}_+$ of the survival cdf, so matching call prices corresponds to matching parts of the integral of the cdf.

3 Construction of the market and model RNDs

To have a very fast RND calibration method, one could choose a model that has closed form expressions for the density (or cdf) of returns. However, in most models used in option pricing, densities (or cdfs) are not known in closed form. As a consequence, it is important to have efficient methods that can approximate them accurately. We present in section 3.1 different numerical techniques to find an approximation $\hat{g}_T$ of the true model density g_T, for any $T > 0$. Then in section 3.2, we explain how

[51] The brackets are zero if $\lim_{s \to \infty} s(1 - \hat{F}_{T_j}(s)) = 0$ or equivalently if $\lim_{s \to \infty} s \int_s^\infty \hat{f}_{T_j}(s) ds = 0$. In particular if $\hat{f}_{T_j}$ has a power tail of the form $\frac{\alpha}{s^\beta}$, it requires that $\beta > 2$.

the mixture of log normal approach can be implemented in order to approximate the market density $\hat{f}_T$.

3.1 Approximating model RNDs

Computing the model distributions of the underlying S_T at some future date $T > 0$ is very similar to computing option prices. On the one hand, it is always possible and rather simple to use Monte Carlo methods and simulate the underlying up to maturity using the Euler scheme (or more advanced schemes[52]) or directly simulate the underlying at maturity and then use the empirical cdf (or a smoothed version) to have an approximation for the model cdf. On the other hand, many option pricing models have closed form characteristic functions for the returns of the underlying S, thus making it possible to find an approximation of the model pdf and cdf via Fourier inversion (assuming the pdf exists). Fourier inversion can be performed in various ways, via FFT, saddlepoint approximations[53], etc. Here we present briefly how we can use Fourier series (used in option pricing as the COS method of Fang and Oosterlee (2008)) to obtain a very fast and accurate approximation of the pdf in practice.

We start by working with the log returns $X_T := \ln(S_T/S_0)$ and we wish to approximate the density $g_{X_T}(x)$ of X_T given $X_0 = 0$ (that we assume to exist). A similar argument can be derived for the cdf. We further assume that the model has a closed form characteristic function:

$$\Psi(u) := \mathbb{E}\left[e^{iuX_T}\right],\ u \in \mathbb{R}.$$

It is the case for many Lévy processes (see Cont and Tankov (2004a)), models of the affine class (see Duffie, Pan, and Singleton (2000)) and many time changed Lévy processes Carr and Wu (2004) making this method applicable to a wide range of models.

The Fourier cosine expansion of the density g_{X_T} can be written as

$$\begin{aligned} g_{X_T}(x) &= \sum_{n=0}^{\infty}{}' A_n \cos\left(n\pi\frac{x-a}{b-a}\right), \\ A_n &= \frac{2}{b-a}\int_a^b g_{X_T}(x)\cos\left(n\pi\frac{x-a}{b-a}\right)dx, \end{aligned} \tag{5}$$

assuming that the density g_{X_T} has finite support $[a, b]$ and where $\sum'$ means that the first term of the sum is divided by two.

Obviously, the support of X_T is in general not finite and to use the cosine expansion, we need to truncate the support of g_{X_T} to a finite interval $[a, b]$ sufficiently wide so that the density is well approximated. It is important to note that the interval $[a, b]$ may depend on the model parameters

[52] See for instance Kloeden and Platen (1992) and Glasserman (2003)
[53] See Aït-Sahalia and Yu (2006), Butler (2007) for instance

since they control the skewness and kurtosis of the distribution g_{X_T}. We refer to Fang and Oosterlee (2008) to choose appropriately the interval $[a, b]$ depending on the cumulants of g_{X_T} and we consider that $[a, b]$ is now fixed. The second step is to replace the infinite sum in (5) by a finite sum. Depending on the regularity of the density, the finite sum converges algebraically or exponentially[54].

As shown in Fang and Oosterlee (2008), the coefficients A_n can be easily approximated using the characteristic function:

$$A_n \approx \frac{2}{b-a} Re\left\{\Psi(\frac{n\pi}{b-a})\exp(-ia\frac{n\pi}{b-a})\right\},$$

3.2 Approximating market RNDs

There is a huge literature on the approximation of the empirical risk-neutral density (alternatively the state price density). Here we advocate the particular use of the mixture of log-normal approach of Alexander and Narayanan (2001), Brigo and Mercurio (2002) because it is able to recover the market RND using only very few quoted option prices. This is important in practice as there may be only one maturity of liquid options quoted and the number of options may be less than ten. For completeness, we briefly present this method, more details can be found in Rebonato (2004), section 9.9.

The starting point is to consider one slice of options' data: $\{C(K_i, T)\}_{1\leq i\leq N}$ for some maturity $T > 0$. In this method, each slice is interpolated/extrapolated independently. We assume we have only call options since all puts can be converted into calls using the put-call parity. If one is using market data, it is important to treat the data and keep only the most informative options. We refer to Aït-Sahalia and Lo (1998), Fan and Mancini (2009) for a details.

The idea is to write the risk-neutral distribution of the market as a mixture of log-normal distributions. We make the following approximation

$$\hat{f}_T(s) = \sum_{k=1}^{p} \alpha_k \varphi(s; \mu_k, \sigma_k),$$

with the constraint that the positive weights $\{\alpha_k\}_{1\leq k\leq p}$ sum to one[55]: $\sum_{k=1}^{p} \alpha_k = 1$. The functions $\varphi(s; \mu_k, \sigma_k)$ are log-normal densities with mean $S_0 e^{\mu_k T}$ and variance $\left(S_0 e^{\mu_k T}\right)^2 \left[e^{\sigma_k^2 T} - 1\right]$. $\mu_k \in \mathbb{R}$ and $\sigma_k \in \mathbb{R}_+^*$ are the parameters of the log normal distributions.

Furthermore, to make sure that $\hat{f}$ is a proper risk-neutral distribution, the following no-arbitrage

[54] In practice this method is extremely accurate for a moderate number of cosine functions (see Fang and Oosterlee (2008)).
[55] This constraint can be easily enforced by changing variables and using trigonometric functions, see Rebonato (2004).

condition needs to be enforced[56]:

$$\sum_{k=1}^{p} \alpha_k e^{\mu_k T} = e^{rT},$$

where r is the risk-free rate of interest.

As a consequence of this setting, the model-implied option price $\hat{C}(K_i, T)$ of the call option $C(K_i, T)$ under the above distributional assumption for $\hat{f}_T$ is

$$\hat{C}(K_i, T) = \sum_{k=1}^{p} \alpha_k C_{BS}(\mu_k, \sigma_k, K_i, T),$$

where $C_{BS}(\mu_k, \sigma_k, K_i, T)$ is the Black-Scholes formula for call options with risk-free rate μ_k, volatility σ_k, strike K_i and maturity T.

We note that the mixture of log-normal distributions is not log-normal and can adapt to the various shapes a density can have. In particular, it can be multi-modal if needed as soon as $p \leq 2$. This can be of interest when capturing the RND generated by jump diffusion models.

To find the parameters of the log-normal mixture $\hat{f}_T$, we will minimize some distance between the market option prices $\{C(K_i, T)\}_{1 \leq i \leq N}$ and the mixture prices $\{\hat{C}(K_i, T)\}_{1 \leq i \leq N}$. We suggest to minimize the least squares distance[57] between prices:

$$d(\hat{C}, C) = \sum_{i=1}^{N} \left(\hat{C}(K_i, T) - C(K_i, T) \right)^2,$$

depending on the parameters $\{\alpha_k, \mu_k, \sigma_k\}_{1 \leq k \leq p}$.

Note that computing the market RNDs as mixture of log-normals only needs to be done once for each slice. This can be considered as pre-processing of data. We suggest a number $p \in \{3, ..., 6\}$ of log-normal distributions in the mixture depending on the structure of the data. We show some examples in section 4.

[56] In practice, we follow Rebonato (2004) and force this condition by computing μ_1 as: $\mu_1 = \frac{1}{T} \ln \left(\frac{e^{rT} - \sum_{k=2}^{p} \alpha_k e^{\mu_k T}}{\alpha_1} \right)$. Although the quantity in the log may be non-positive, it has never been the case in practice.

[57] Given that we are interested in fitting model and market RNDs and we want to match prices, it makes sense to consider a loss function based on prices (and not implied volatilities for instance). Nevertheless, in practice it does not matter so much which loss function is considered because the approximating prices $\{\hat{C}(K_i, T)\}_{1 \leq i \leq N}$ are almost exact (up to 8 digits) and the resulting mixture are almost the same. We find that minimizing squared prices leads to the fastest minimization and the least local minima.

4 Results and performance testing

As we want to evaluate the performance of the RND calibration method without our results being influenced by model misspecification, we choose to work with simulated data. In particular, we illustrate our results in the case of the Heston model (Heston (1993)). This model has a closed form characteristic function so Fourier methods can be used to compute the RND. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{Q})$ be a filtered probability space satisfying the usual assumptions, where $\mathbb{Q}$ denotes a risk-neutral measure. We recall the dynamics of returns in the Heston model:

$$\begin{aligned}
&\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dW_t, \\
&dv_t = \kappa(\theta - v_t)dt + \sigma_v\sqrt{v_t}dZ_t, \\
&S_0, v_0 > 0, \quad \text{and } d[W,Z]_t = \rho dt,
\end{aligned}$$

where κ is the speed of mean reversion of the variance process v, θ is its level of mean reversion and σ_v is the volatility of the variance, sometimes referred to as the volvol parameter. We assume that the Feller condition $2\kappa\theta > \sigma_v^2$ is enforced so that the variance process v remains almost surely positive. The coefficient ρ denotes the correlation between changes in returns and changes in their variance, and makes it possible to represent the so-called leverage effect. We refer to Heston (1993) for more details.

Our testing procedure goes as follows. First, we randomly pick model parameters[58] and generate a certain number N of option prices. To make the identification of the parameters potentially more difficult and test the calibration techniques under difficult conditions, we choose N as low as $N = 5$ and only one maturity T. We find that when the number of strikes is large, the market RND is very well identified and as a result the RND and option price calibrations yield very similar results. We therefore report the results for $N = 5$. We present the results for uniformly spaced strikes, as it turn out that results do not differ qualitatively when simulated strikes are more dense around the initial stock price[59] S_0 (the extreme strikes being the same). The second step consists in approximating the market RND $\hat{f}_T$ and we detail the quality of this approximation in section 4.1. Finally, we randomly generate model parameters that we use as starting point to minimize the L^1 distance between the model and market survival cdfs with an emphasis on the part of the distribution we have data on (in the range of strikes traded). We motivate this choice of loss function and give the results for in and out of pricing in section 4.2.

In all tests, we consider the risk-free interest rate to be $r = 1\%$ and the initial stock price to be $S_0 = 50$. Additionally, even though we have tested the RND calibration for various parameter values, we expose the results for only two different sets of parameters as they are representative of our overall

[58] We nevertheless make sure that the Feller condition is enforced and that simulated parameters can generate deep out of the money option prices that are larger than 0.01. When simulating parameters blindly, it is often the case that option prices are smaller than 1e-4 for strikes $K < 0.70S_0$ for instance.

[59] In practice, more strikes are traded at the money, i.e. for strikes close to today's stock price (or Futures price).

results. We consider:

$$\text{Param1} : \kappa = 4.2; \sigma_v = 0.80; \rho = -0.85; v_0 = 0.02; \theta = 0.08, \tag{6}$$

$$\text{Param2} : \kappa = 3; \sigma_v = 0.40; \rho = -0.20; v_0 = 0.06; \theta = 0.03. \tag{7}$$

These parameters correspond to different market scenarios as one can see on Figure 18, where we have plotted the corresponding implied volatility smiles. While the first set of parameters generates a smile which is linear in the strike, the second set produces a smile which is very steep. Out of the money puts are therefore much more expensive in the second scenario than in the first one, which is usually the case in a volatile market, when market participants believe that a large movement may happen before maturity.

We simulate for Param1 and Param2 $N = 5$ strikes uniformly distributed in the spot-moneyness space $m = K/S_0$ with moneyness between $m = 0.9$ and $m = 1.05$ (this corresponds to strikes between $K = 45$ and $K = 52.5$) which means that traded options have strikes that are close to at-the-money with slightly more puts traded than calls traded. This corresponds to a realistic setting, especially for individual stocks where only few options are traded. In the following, we will assume that the maturity considered for Param1 is $T = 6/12$ and $T = 1/12$ for Param2, unless otherwise stated.

4.1 Accuracy of the mixture of log-normal approach

We fit the mixture of log-normals to option prices using the method described in section 3.2 and minimizing[60] the squared dollar error between option prices. It is necessary to have between three and four log-normal densities to fit option prices perfectly. Param1 (resp. Param2) requires a mixture of $p = 3$ (resp. $p = 4$) log-normals to achieve a sum of squared pricing error of 1e-16.

Knowing the true parameters of the model, one can compute the market (true) densities f_T (and cdf F_T) and the mixture pdf $\hat{f}_T$ (and cdf $\hat{F}_T$). We show in Figure 19 the RND fits when only 5 strikes are available between $K = 45$ and $K = 52.5$. We immediately see that having only five option prices can lead to a rather crude approximation (for Param1) or a very good approximation (Param2). In particular, for both parameter sets, it is rather surprising that the approximation is so good considering that only strikes between $K = 45$ and $K = 52.5$ were given as inputs. The quality of the fit depends on the data generating process. In fact increasing the number of log-normals in the mixture does not improve the fit for Param1, however having a couple more strikes (say $N = 10$) especially for smaller moneyness, enables to recover the market RND almost perfectly. We want to stress that despite an seemingly bad fit for Param1, $\sup_{\mathbb{R}_+} |f_T - \hat{f}_T|$ is of the order of 1e-3 both for Param1 and Param2 (although the relative error is larger for Param1).

We do not display the resulting fit for the approximation $\hat{F}_T$ of the market cdf F_T, as it would not

[60] To cope with the ill-posedness of the mixture calibration problem and the potential existence of multiple minima, we use a global optimizer namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996). We are grateful to Jochen Krause for his implementation of various evolution optimizers including the CMA-ES algorithm. Finding an optimal solution only takes a few seconds on a single core 2.2GHz.

bring additional information but it is interesting to note that the error for the infinity norm is only slightly larger than for the densities. In the interval where strikes are available the approximation for the cdf is very accurate as we also have an infinity norm of the order of 1e-3. The relative errors on the cdfs are therefore smaller than for the densities. As a consequence we suggest to use the estimator $\hat{F}_T$ for calibration purposes.

4.2 Calibration results

4.2.1 Choice of loss function for the RND calibration

Given the results on the quality of the approximations $\hat{f}_T$ and $\hat{F}_T$, we choose a loss function for the RND calibration based on the cdf[61]. Given the high uncertainty in the tail extrapolation, we also recommend to minimize distance (2) for a weighting function ω that simply truncates the area where no strikes are traded. We choose to minimize the L^1 distance between (survival) cdfs as we know that they are integrable. Moreover, denoting K_1 and K_N the smallest and largest strikes quoted, we have chosen in our numerical experiments[62]:

$$\omega(ds, dt) = \delta_T(dt) \frac{1}{K_N - K_1} 1_{[K_1, K_N]}(s) ds,$$

which leads us to minimize

$$d^1_\omega(\hat{\bar{F}}_T, \bar{G}_T) = \frac{1}{K_N - K_1} \int\limits_{s \in [K_1, K_N]} \left| \hat{F}_T(s) - G_T(s) \right| ds. \tag{8}$$

We then discretize this integral using an equidistant grid on $[K_1, K_N]$.

4.2.2 In and out of sample results

We generate random parameters for the model that we use as starting values in the optimization. We then find two sets of optimal parameters, one minimizing the distance (8), and the other one minimizing the least squares option pricing errors:

$$d^2_{\mathrm{LS}}(\tilde{C}, C) = \sum_{i=1}^{N} \left(\tilde{C}(K_i, T) - C(K_i, T) \right)^2, \tag{9}$$

where $\tilde{C}(K_i, T)$ denotes the model option price for strike K_i and maturity T.

As expected, we find that minimizing both loss functions takes approximately the same amount

[61] Additionally, we do not recommend the use of relative entropy measures when $\hat{f}_T$ is a mixture of log-normals since this distance criterion emphasizes the fitting of the tails and the mixture eventually has tails of the log-normal type (even if the body may be very different from a log-normal distribution). In any case, emphasizing the fit far in the tails where no data is available and the RND is extrapolated makes little sense regardless of the assumption of mixture of log-normals.
[62] Truncating less can sometimes give better out of sample results, but this is not systematic.

of time. However, the parameters found by RND calibration are closer (in terms of relative errors) to the true parameters than those found by least squares calibration and this is especially the case when there are only a few strikes traded. It therefore seems that using a RND calibration can help to better identify the true generating process because of the additional structure given as input to the calibration. This remark also holds in the case where the approximate mixture density $\hat{f}_T$ is wiggly as for Param1 (see Figure 19).

As expected, the least squares calibration to option prices leads to better in sample fitting of option prices, i.e. smaller pricing errors. Interestingly for Param2, the fit of option prices via RND calibration is as good as with the least squares calibration. This is not the case for Param1. These results can be explained by the fact that the distribution $\hat{F}_T$ for Param2 is very close to the true distribution F_T which is attainable by the model since it is well specified. In the case of Param1, $\hat{F}_T$ is not close to any distribution that can be generated via the model and the RND calibration leads to worse option price estimates.

In Figure 20, we show how the optimal parameters found via RND or least squares calibration perform. We have plotted for both Param1 and Param2 the true implied volatilities (IVs) as well as the RND IVs (IVs generated by the optimal parameters obtained via RND calibration) and the least squares IVs (IVs generated by the optimal parameters obtained via least squares calibration). The calibration was performed for ranges of strikes in $[45, 52.5]$ and we show the out of sample results for strikes down to $K = 17.5$ and as high as $K = 65$ (which correspond to moneynesses $m = K/S_0$ in the range $[0.35, 1.30]$). The plot confirms that in the case of Param1, the target density $\hat{f}_T$ was not attainable by the model and we see that there is a irreducible distance between the RND IVs and the true IVs except for very extreme strikes. This is not the case for Param2 where the RND and least squares IVs match perfectly the true IVs. The least squares IV fits tend to deteriorate outside the range of strikes we calibrated the model to but this is not the case for the RND IVs that often get closer to the true IVs as the strikes become extremely small or large. We attribute this interesting outcome to the fact that RND calibration adds structure to the existing data and provided that this structure is well specified, it adds identifiability to the calibration problem.

Appendix

1 Figures

Figure 18: Implied volatilities for the set of parameters Param1 and Param2. The spot price is $S_0 = 50$.

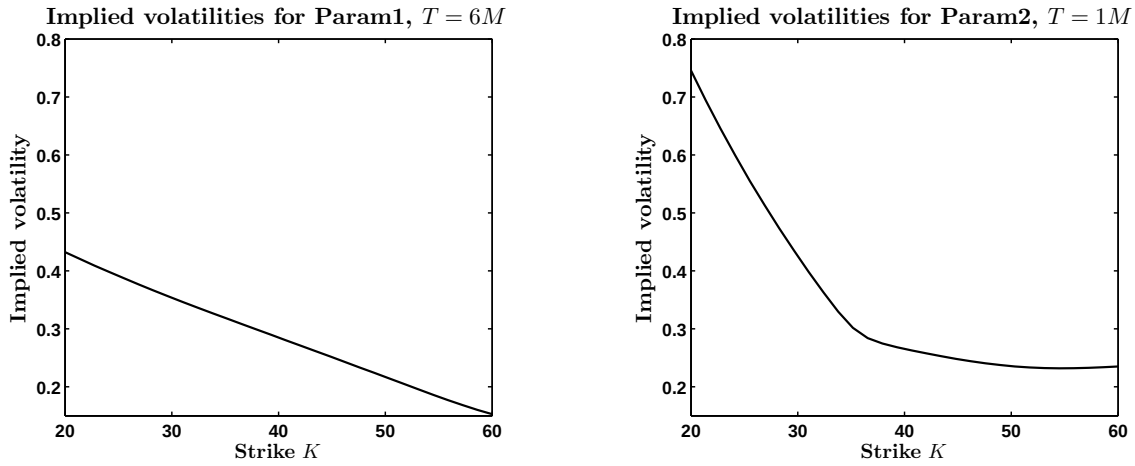


Figure 19: True density f_T and approximation $\hat{f}_T$ for the model parameters Param1 and Param2. The mixtures are computed using respectively $p = 3$ and $p = 4$ log-normals given that there are only 5 strikes available on the market between $K = 45$ and $K = 52.5$. The spot price is $S_0 = 50$.

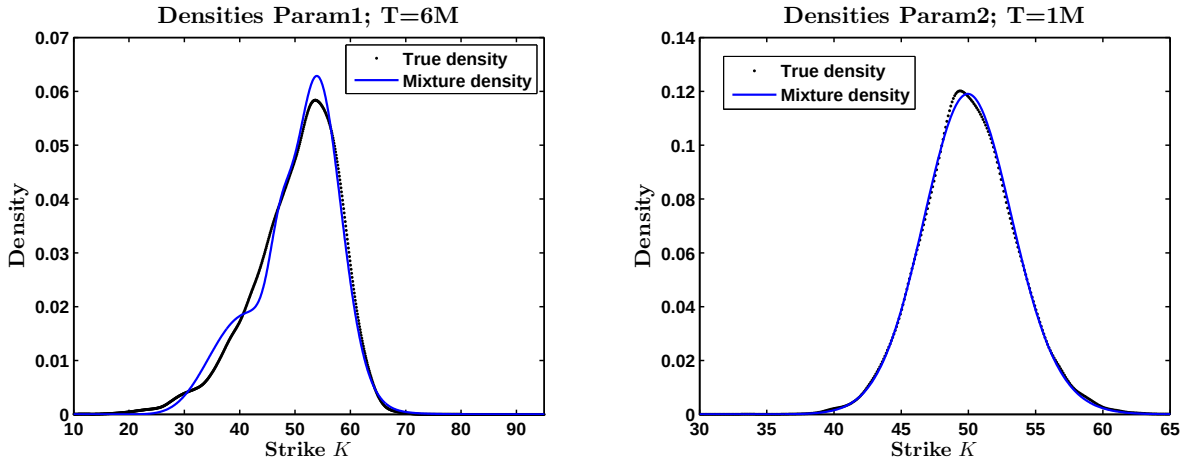
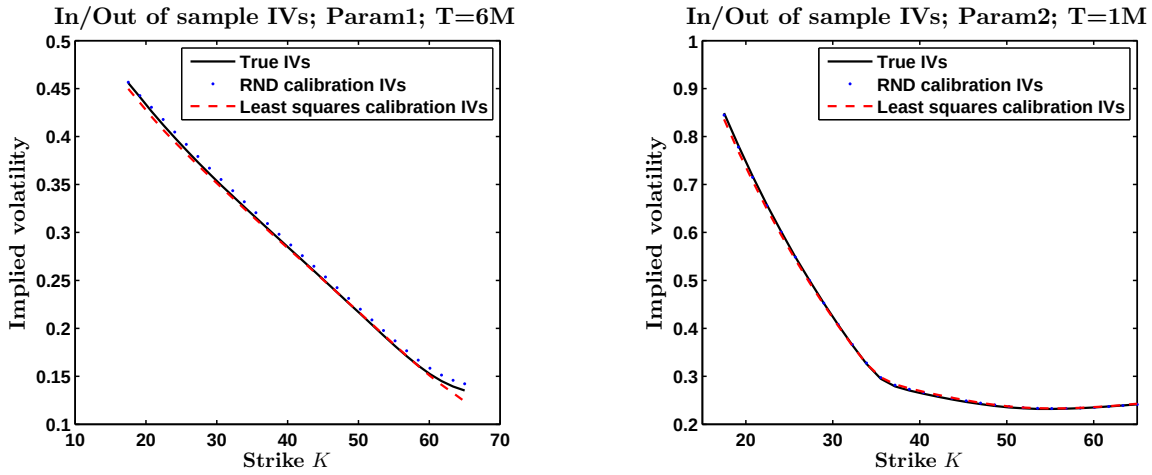


Figure 20: Implied volatilities generated by the optimal parameters found after a RND calibration and a least squares optimization (to option prices). The spot price is $S_0 = 50$ and both calibrations were done to only 5 strikes in the range $[45, 52.5]$.



REFERENCES

Aït-Sahalia, Y., and J. Duarte, 2003, "Nonparametric option pricing under shape restrictions," *The Journal of Econometrics*, 116, 9 –47.

Aït-Sahalia, Y., and A. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53, 500 –547.

Aït-Sahalia, Y., and J. Yu, 2006, "Saddlepoint approximations for continuous-time Markov processes," *Journal of Econometrics*, 134, 507–551.

Alexander, C., and S. Narayanan, 2001, "Option Pricing with Normal Mixture Returns: Modelling Excess Kurtosis and Uncertanity in Volatility," Icma centre discussion papers in finance.

Andersen, L., and R. Brotherton-Ratcliffe, 1998, "The equity option volatility smile: an implicit finite-difference approach," *Journal of Computational Finance*, 1(2).

Avellaneda, M., 1998, "Minimum-Relative-Entropy Calibration of Asset-Pricing Models," *International Journal of Theoretical and Applied Finance*, 1(4), 447–472.

Bahra, B., 1997, "Implied risk-neutral probability density functions from option prices: theory and application," Bank of england working papers, Bank of England.

Bakshi, G., C. Cao, and Z. Chen, 1997, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–2049.

Bentata, A., and R. Cont, 2009, "Mimicking the marginal distributions of a semimartingale," .

Black, F., and M. Scholes, 1973, "The pricing of options and corporate liabilities," *Journal of Political Economy*, 81, 637–654.

Bliss, R. R., and N. Panigirtzoglou, 2004, "Option-Implied Risk Aversion Estimates," *The Journal of Finance*, 59(1), pp. 407–446.

Breeden, D., and R. Litzenberger, 1978, "Prices of state-contingent claims implicit in option prices," *Journal of Business*, 51, 621 –651.

Brigo, D., and F. Mercurio, 2002, "Lognormal-mixture dynamics and calibration to market volatility smiles," *International Journal of Theoretical and Applied Finance*, 05(04), 427–446.

Bu, R., and K. Hadri, 2007, "Estimating option implied risk-neutral densities using spline and hypergeometric functions," *Econometrics Journal*, 10(2), 216–244.

Butler, R. W., 2007, *Saddlepoint approximations with applications*. Cambridge University Press.

Carr, P., and L. Wu, 2004, "Time-changed Levy processes and option pricing," *Journal of Financial Economics*, 71(1), 113–141.

Christoffersen, P., and K. Jacobs, 2004, "The Importance of the Loss Function in Option Valuation," *Journal of Financial Economics*, 72, 291–318.

Cont, R., and P. Tankov, 2004a, "Non-parametric calibration of jump-diffusion option pricing models," *Journal of Computational Finance*, 7, 1–49.

——— , 2004b, "Nonparametric calibration of jump-diffusion option pricing models," *Journal of Computational Finance*, 7, 1–49.

——— , 2006, "Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem," *SIAM J. Control Optim*, pp. 1–25.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Dupire, B., 1994, "Pricing with a Smile," *Risk*, 7, 18–20.

Fan, J., and L. Mancini, 2009, "Option Pricing with Model-Guided Nonparametric Methods," *Journal of the American Statistical Association*, forthcoming.

Fang, F., and C. W. Oosterlee, 2008, "A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions.," *SIAM J. Scientific Computing*, 31(2), 826–848.

Glasserman, P., 2003, *Monte Carlo Methods in Financial Engineering*. Springer.

Gourieroux, C., J. Jasiak, and R. Sufana, 2009, "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150(2), 167 – 181.

Gyöngy, I., 1986, "Mimicking the one-dimensional marginal distributions of processes having an ito differential," *Probability Theory and Related Fields*, 71(4), 501–516.

Hansen, N., and A. Ostermeier, 1996, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC '96)*, pp. 312–317.

Heston, S., 1993, "A closed-form solution for options with stochastic volatility, with applications to bond and currency options," *Review of Financial Studies*, 6, 327–343.

Heston, S. L., and S. Nandi, 2000, "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13(3), 585–625.

Jackwerth, J., and M. Rubinstein, 1996, "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51(5), 1611–1631.

Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, Germany.

Madan, D., P. Carr, and E. Chang, 1998, "The Variance Gamma process and option pricing," *European Finance Review*, 2, 79–105.

Madan, D., and M. Yor, 2002, "Making Markov martingales meet marginals: with explicit constructions," *Bernoulli*, 8(4), 509–536.

Rebonato, R., 2004, *Volatility and correlation : the perfect hedger and the fox*. John Wiley & Sons, Inc., 2nd. edn.

Ziegler, A., 2007, "Why Does Implied Risk Aversion Smile?," *The Review of Financial Studies*, 20(3), 859–904.

Part III

Appendix

Chris Bardgett - Curriculum Vitae

Date of birth: 21.10.1983

Education

Sep. 2008– July 2014	**PhD in Banking and Finance** at the Swiss Finance Institute and University of Zürich, Switzerland
Jan. 2013– July. 2013	**Visiting Research Fellow** at Imperial College London (Mathematical Finance section).
Sep. 2006– Mar. 2008	**Master of Quantitative Finance** at ETH and University of Zurich Master thesis: *Pricing convertible bonds using Finite Elements*
Sep. 2004– Sept. 2008	Student at ENSIMAG, the **French Engineering School of Computer Sciences and Applied Mathematics** in Grenoble.

Work experience

Sep. 2007– Aug. 2008	**Quantitative Analyst at UBS** (Zürich) Quantitative Risk Control Group (model validation team). Work on the modeling and implementation (C/C++/Matlab) of convertible bonds for pricing and risk management. Numerical treatment of free boundary problems and PDEs using Finite Elements.
Summer 2013	**Intern at Natixis** (a French investment bank). Work with traders to implement a software for pricing and managing risks of capital guaranteed products (CPPI).